

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 21:41:43 ; Search time 2063 Seconds  
(without alignments)  
2721.442 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtggggtcgagctaagca.....ttcaccatgaggtgtgtt 231

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_esthu.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_bhg.\*  
27: em\_gss\_vri.\*  
28: gb\_gss1.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	231	100.0	477	28	AZ231159	AZ231159 RPCI-23-5
2	231	100.0	501	28	A2440299	A2440299 1M0231K17
3	231	100.0	527	10	BG087407	BG087407 H3139B12-
4	231	100.0	558	28	BH036520	BH036520 RPCI-24-2

5	231	100.0	559	28	BH105759	BH105759 RPCI-24-2
6	231	100.0	577	28	AZ420721	AZ420721 1M0198L07
c	7	231	587	28	A2832364	A2832364 2M0112B13
c	8	231	627	28	A2495883	A2495883 1M0332E03
c	9	231	647	28	A2993990	A2993990 2M0279D16
c	10	231	706	14	C8056516	C8056516 NISC-J117
c	11	231	721	28	A2994818	A2994818 2M0280J24
c	12	229.4	472	28	A2310386	A2310386 1M0025O13
c	13	229.4	505	28	A2054649	A2054649 RPCI-23-4
c	14	229.4	518	9	AW558986	AW558986 L0302F10-
c	15	229.4	520	12	BM219777	BM219777 C0929H06-
c	16	229.4	534	28	A2341190	A2341190 1M0073M24
c	17	229.4	535	28	A2326140	A2326140 1M0048D11
c	18	229.4	539	28	A2965228	A2965228 2M0235A17
c	19	229.4	557	9	AW554086	AW554086 L0235F05-
c	20	229.4	558	28	A2987016	A2987016 2M0269M23
c	21	229.4	584	28	A2432413	A2432413 1M0217N15
c	22	229.4	588	28	A242504	A242504 RPCI-23-8
c	23	229.4	662	28	A2333042	A2333042 1M0061J16
c	24	229.4	678	14	BY731717	BY731717 BY731717
c	25	229.4	686	28	BH092157	BH092157 RPCI-24-2
c	26	229.4	707	28	BH117452	BH117452 RPCI-24-2
c	27	229.4	834	28	BH062187	BH062187 RPCI-24-3
c	28	229.4	4484	11	AK033082	AK033082 Mus_muscu
c	29	228.4	606	12	B1134544	B1134544 UI-M-BH3-
c	30	227.8	309	28	BH068706	BH068706 RPCI-24-3
c	31	227.8	457	28	A2071850	A2071850 RPCI-23-4
c	32	227.8	492	9	AW552509	AW552509 L0213E08-
c	33	227.8	531	28	A2636367	A2636367 1M0495J10
c	34	227.8	557	28	BH078165	BH078165 RPCI-24-2
c	35	227.8	559	13	BQ553681	BQ553681 H4033F02-
c	36	227.8	596	28	A2830822	A2830822 2M0110H14
c	37	227.8	646	28	BH099567	BH099567 RPCI-24-3
c	38	227.8	678	28	A2897416	A2897416 RPCI-24-2
c	39	227.8	686	28	A2259805	A2259805 RPCI-23-1
c	40	227.8	706	28	A2797045	A2797045 2M0053J06
c	41	227.8	793	28	A2719135	A2719135 RPCI-24-1
c	42	227.8	838	28	BH027115	BH027115 RPCI-24-2
c	43	227.4	553	28	A2792711	A2792711 2M0045C19
c	44	227.4	642	28	A2090071	A2090071 RPCI-23-4
c	45	227.4	675	28	A2957637	A2957637 2M0224N06

## ALIGNMENTS

RESULT 1  
AZ231159

LOCUS  
RPCI-23-51K23.TJ

DEFINITION  
genomic survey sequence.

ACCESSION  
AZ231159

VERSION  
AZ231159.1

KEYWORDS  
GSS.

SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 477)

AUTHORS  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aklnrst, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.

TITLE  
Mouse BAC End Sequences from Library RPCI-23

JOURNAL  
Unpublished

COMMENT  
Other GSSs: RPCI-23-51K23.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@igr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC

AZ231159 477 bp DNA linear GSS 14-JUN-2000  
RPCI-23-51K23.TJ RPCI-23 Mus musculus genomic clone RPCI-23-51K23,  
genomic survey sequence.

AZ231159 GI:8539205

GSS.

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 477)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aklnrst,

B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished

Other GSSs: RPCI-23-51K23.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@igr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 51 row: K column: 23  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source

Location/Qualifiers  
 1. .477  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-51K23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /clone\_lib="RPCI-23"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 94 a 110 c 143 g 130 t

BASE COUNT 94 a 110 c 143 g 130 t  
 ORIGIN  
 Query Match 100.0%; Score 231; DB 28; Length 477;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-60;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGCTAAGCACTACACAGAGATAGCTGTGTGGCATCTCTGGAAGG 60  
 |||||  
 Db 232 GTGGGTGCGAGCTAAGCACTACACAGAGATAGCTGTGTGGCATCTCTGGAAGG 291  
 QY 61 CAGCTCTGATGATGAGTTCAGTGTCTAGTCTCCCTCCCGAGAAAACGACACG 120  
 |||||  
 Db 292 CAGCTCTGATGATGAGTTCAGTGTCTAGTCTCCCTCCCGAGAAAACGACACG 351  
 QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAAGGGATGTTTGTAGGGCCCT 180  
 |||||  
 Db 352 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAAGGGATGTTTGTAGGGCCCT 411  
 QY 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTCAACCATGAGGCTTGCTT 231  
 |||||  
 Db 412 ATGCTTGACACATGGGGATCAGACCTCTACCTTCAACCATGAGGCTTGCTT 462

RESULT 2  
 AZ440299 501 bp DNA linear GSS 03-OCT-2000  
 LOCUS 1M0231K17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC1M0231K17 F, genomic survey sequence.

ACCESSION AZ440299  
 VERSION AZ440299.1 GI:10564312  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0231 row: K column: 17  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 501.  
 Location/Qualifiers  
 1. .501

FEATURES  
 source

/organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0231K17"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gill4732114[gb]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 119 a 123 c 136 g 123 t  
 ORIGIN  
 Query Match 100.0%; Score 231; DB 28; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-60;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGCTAAGCACTACACAGAGATAGCTGTGTGGCATCTCTGGAAGG 60  
 |||||  
 Db 94 GTGGGTGCGAGCTAAGCACTACACAGAGATAGCTGTGTGGCATCTCTGGAAGG 153  
 QY 61 CAGCTCTGATGATGAGTTCAGTGTCTAGTCTCCCTCCCGAGAAAACGACACG 120  
 |||||  
 Db 154 CAGCTCTGATGATGAGTTCAGTGTCTAGTCTCCCTCCCGAGAAAACGACACG 213  
 QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAAGGGATGTTTGTAGGGCCCT 180  
 |||||  
 Db 214 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAAGGGATGTTTGTAGGGCCCT 273  
 QY 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTCAACCATGAGGCTTGCTT 231  
 |||||  
 Db 274 ATGCTTGACACATGGGGATCAGACCTCTACCTTCAACCATGAGGCTTGCTT 324

RESULT 3  
 BG087407 527 bp mRNA linear EST 13-JUN-2003  
 LOCUS H3139B12-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
 DEFINITION H3139B12 5', mRNA sequence.  
 ACCESSION BG087407  
 VERSION BG087407.1 GI:12569971  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
MEDLINE  
PUBMED  
COMMENT

Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 527)  
Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac  
,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.  
III, Becker,K.G. and Ko,M.S.H.  
Genome-wide expression profiling of mid-gestation placenta and  
embryo using a 15,000 mouse developmental cDNA microarray  
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
20381348  
10922068  
Other\_ESTs: H3139B12-3  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit <http://igsun.grc.nia.nih.gov/cdna/15k.html> for details.  
Plate: H3139 row: B column: 12  
Seq primer: -21M13 Reverse  
High quality sequence stop: 527  
POLYA=NO.

## FEATURES

source

Location/Qualifiers

1. .527  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="niaEST:H3139B12-5"  
/db\_xref="taxon:10090"  
/clone="H3139B12"  
/sex="Clones arrayed from a variety of cDNA libraries"  
/dev\_stage="Clones arrayed from a variety of cDNA libraries"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse 15K cDNA Clone Set"  
/note="Vector: pSPORT1; Site.1: SalI; Site.2: NotI; This  
clone is among a rearranged set of 15,247 clones from 11  
embryo cDNA libraries (including preimplantation stage  
embryos from unfertilized egg to blastocyst, embryonic  
part of E7.5 embryos, extraembryonic part of E7.5 embryos  
, and E12.5 female mesonephros/gonad) and one newborn  
ovary cDNA library. Average insert size 1.5 kb. All  
source libraries are cloned unidirectionally with Oligo(dT  
)-Not primers. References include: (1) Genome-wide  
expression profiling of mid-gestation placenta and embryo  
using a 15,000 mouse developmental cDNA microarray, 2000,  
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)  
Large-scale cDNA analysis reveals phased gene expression  
patterns during preimplantation mouse development, 2000,  
Development, 127: 1737-1749; (3) Genome-wide mapping of  
unselected transcripts from extraembryonic tissue of  
7.5-day mouse embryos reveals enrichment in the t-complex  
and under-representation on the X chromosome, 1998, Hum  
Mol Genet 7: 1967-1978."

BASE COUNT 104 a 121 c 138 g 144 t

ORIGIN

Query Match 100.0%; Score 231; DB 10; Length 527;  
Best Local Similarity 100.0%; Pred. No. 1.6e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCAGCTGCACAGAGGATAGCTTCTGTGGCATCTCTGTGAAGG 60  
|||||  
DB 236 GTGGGTGGAGGCTAAGCAGCTGCACAGAGGATAGCTTCTGTGGCATCTCTGTGAAGG 295  
|||||  
QY 61 CAGCTCTGATTCGATGAAGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAACGACAG 120  
|||||  
DB 296 CAGCTCTGATTCGATGAAGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAACGACAG 355  
|||||  
QY 121 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
|||||  
DB 356 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 415  
|||||

QY 181 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
|||||  
DB 416 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 466  
|||||

## RESULT 4

BH036520/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH036520  
RPCI-24-254C6.TJ RPCI-24 Mus musculus genomic clone RPCI-24-254C6,  
genomic survey sequence.  
BH036520  
BH036520.1 GI:14811469  
GSS.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 558)  
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,  
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,  
Russell,D., de Jong,P. and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished  
Other\_GSSs: RPCI-24-254C6.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 254 row: C column: 6  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers

1. .558  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-254C6"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/clone\_lib="RPCI-24"  
/note="Vector: pTARBAC1; Site.1: BamHI; Site.2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 151 a 158 c 129 g 120 t

ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 558;  
Best Local Similarity 100.0%; Pred. No. 1.6e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCAGCTGCACAGAGGATAGCTTCTGTGGCATCTCTGTGAAGG 317  
|||||  
DB 376 GTGGGTGGAGGCTAAGCAGCTGCACAGAGGATAGCTTCTGTGGCATCTCTGTGAAGG 317  
|||||  
QY 61 CAGCTCTGATTCGATGAAGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAACGACAG 120  
|||||  
DB 316 CAGCTCTGATTCGATGAAGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAACGACAG 257  
|||||  
QY 121 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
|||||  
DB 256 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 197  
|||||

```

181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
|||||
Db 196 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 146

RESULT 5
BH105759
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

559 bp DNA linear GSS 19-JUL-2001
RPCI-24-237C16.TJ RPCI-24 Mus musculus genomic clone RPCI-24-237C16
, genomic survey sequence.
BH105759
BH105759.1 GI:14935045
GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 559)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsedaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 237 row: C column: 16
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..559
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-237C16"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/notice="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."
119 a 132 c 159 g 149 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 231; DB 28; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.6e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGGTCCGAGGCTAAGCACTGCACAGATAGCTTGTGTGGCATCTCTGTGAAGG 60
|||||
Db 243 GTGGGGTCCGAGGCTAAGCACTGCACAGATAGCTTGTGTGGCATCTCTGTGAAGG 302
|||||
QY 61 CACGCTCTATTGCATGAAGTTTCAGTTCCTAGTTCCTTCCCTCCCGAGGAAAACGACAG 120
|||||
Db 303 CACGCTCTATTGCATGAAGTTTCAGTTCCTAGTTCCTTCCCTCCCGAGGAAAACGACAG 362
|||||
QY 121 GGAGCTGCCAAGACCTCTCTGGCTGATGACCTTAAGGGATGTTTGTGTAGGCCCTT 180
|||||
Db 363 GGAGCTGCCAAGACCTCTCTGGCTGATGACCTTAAGGGATGTTTGTGTAGGCCCTT 422
|||||
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231

```

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCCTGTGGAAGG 60  
 |||||  
 Db 374 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCCTGTGGAAGG 315  
 |||||  
 QY 61 CAGCTCTGATGCATGAAGGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACACGACG 120  
 |||||  
 Db 314 CAGCTCTGATGCATGAAGGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACACGACG 255  
 |||||  
 QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
 |||||  
 Db 254 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 195  
 |||||  
 QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
 |||||  
 Db 194 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 144  
 |||||

RESULT 7  
 AZ832364/c  
 LOCUS 587 bp DNA linear GSS 20-FEB-2001  
 DEFINITION 2M0112B13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M0112B13 R, genomic survey sequence.  
 AZ832364  
 VERSION  
 ACCESSION AZ832364.1 GI:13002272  
 KEYWORDS  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 587)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0112 row: B column: 13  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 587.  
 Location/Qualifiers  
 1..587  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0112B13"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (g14732114|gb|AF129072.1), a copy-number

Inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 159 a 168 c 135 g 125 t  
 ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 587;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-60;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCCTGTGGAAGG 60  
 |||||  
 Db 246 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCCTGTGGAAGG 187  
 |||||

QY 61 CAGCTCTGATGCATGAAGGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACACGACG 120  
 |||||  
 Db 186 CAGCTCTGATGCATGAAGGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACACGACG 127  
 |||||

QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
 |||||  
 Db 126 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 67  
 |||||

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
 |||||  
 Db 66 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 16  
 |||||

RESULT 8  
 AZ495883/c  
 LOCUS 627 bp DNA linear GSS 05-OCT-2000  
 DEFINITION 1M0332E03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0332E03 F, genomic survey sequence.  
 AZ495883  
 VERSION  
 ACCESSION AZ495883.1 GI:10671633  
 KEYWORDS  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 627)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0332 row: E column: 03  
 Seq primer: CGTTGTAAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 627.  
 Location/Qualifiers  
 1..627  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0332E03"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"

FEATURES  
 source

/note="vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 160 a 190 c 142 g 135 t  
ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 627;  
Best Local Similarity 100.0%; Pred. No. 1.7e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGTGAAGG 60  
Db 441 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGTGAAGG 382

Qy 61 CAGCTCTGATGATGAGGTTGAGTCTAGTCTAGTCTCCCTCCCGGAGGAGGAGGAGG 120  
Db 381 CAGCTCTGATGATGAGGTTGAGTCTAGTCTAGTCTCCCTCCCGGAGGAGGAGGAGG 322

Qy 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTTAAAGGATGTTTGTGTAGGGCCCT 180  
Db 321 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTTAAAGGATGTTTGTGTAGGGCCCT 262

Qy 181 ATGCTTGCACATGGGGATCAGACCTTACCTTACCCATGAGGCTTGCTT 231  
Db 261 ATGCTTGCACATGGGGATCAGACCTTACCTTACCCATGAGGCTTGCTT 211

RESULT 9  
AZ993990  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ993990 647 bp DNA linear GSS 27-APR-2001  
2M0279D16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
clone UUGC2M0279D16 F, genomic survey sequence.

AZ993990  
A2993990.1 GI:13865217  
GSS.  
Mus musculus  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 647)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0279 row: D column: 16

Seq primer: CGTTGTAAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 647.  
Location/Qualifiers  
source

1..647  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0279D16"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/note="vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gii4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 149 a 146 c 185 g 167 t  
ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 647;  
Best Local Similarity 100.0%; Pred. No. 1.7e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGTGAAGG 60  
Db 235 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGTGAAGG 294

Qy 61 CAGCTCTGATGATGAGGTTGAGTCTAGTCTAGTCTCCCTCCCGGAGGAGGAGGAGG 120  
Db 295 CAGCTCTGATGATGAGGTTGAGTCTAGTCTAGTCTCCCTCCCGGAGGAGGAGGAGG 354

Qy 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTTAAAGGATGTTTGTGTAGGGCCCT 180  
Db 355 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTTAAAGGATGTTTGTGTAGGGCCCT 414

Qy 181 ATGCTTGCACATGGGGATCAGACCTTACCTTACCCATGAGGCTTGCTT 231  
Db 415 ATGCTTGCACATGGGGATCAGACCTTACCTTACCCATGAGGCTTGCTT 465

RESULT 10  
CB056516/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

CB056516 706 bp mRNA linear EST 17-JAN-2003  
NISC-jj37f07.w1 Soares NMBP13-15 Mus musculus cDNA clone  
IMAGE:4848516 5', mRNA sequence.  
CB056516  
CB056516.1 GI:27794803  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 706)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cyabs@email.nih.gov  
CDNA Library Preparation:  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
info@image.llnl.gov  
Plate: LLAM10779 row: K column: 13  
Seq primer: T7 primer.

## FEATURES

source

Location/Qualifiers  
1..706  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4848516"  
/tissue\_type="pituitary gland"  
/dev\_stage="juvenile, 13-15 days"  
/lab\_host="DR10B (phage-resistant)"  
/clone\_lib="Soares NMBP13-15"  
/notes="Organ: brain; Vector: pT7T3D-Paci; Site\_1: NotI;  
Site\_2: EcoRI; 1st strand cDNA was primed with a NotI-  
oligo(dT) primer  
5'-ACTGCAAGATTGCGCGCGGTACCGATCTTTTCTTTTCTTTT-3';  
double-stranded cDNA was ligated to EcoRI adaptors  
5'-AATTCGCGAGG-3' and 5'-CTGTGCGG-3' (Pharmacia),  
digested with NotI and cloned into the NotI and EcoRI  
sites of the pT7T3D-Paci vector. Library went through one  
round of normalization, and was constructed in the  
laboratory of M. Bento Soares (University of Iowa)."

BASE COUNT 157 a 199 c 172 g 178 t  
ORIGIN

Query Match 100.0%; Score 231; DB 14; Length 706;  
Best Local Similarity 100.0%; Pred. No. 1.8e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCTCTGTGGAAGG 60  
|||||  
DB 634 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCTCTGTGGAAGG 575  
|||||

QY 61 CAGCTCTGATGCATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAAACGACAG 120  
|||||

DB 574 CAGCTCTGATGCATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAAACGACAG 515  
|||||

QY 121 GGAGCTGGCGACAGACCTCTCTGGGTGATGAGCTTAAGGATGTTTGTGTAGGGCCCT 180  
|||||

DB 514 GGAGCTGGCGACAGACCTCTCTGGGTGATGAGCTTAAGGATGTTTGTGTAGGGCCCT 455  
|||||

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
|||||

DB 454 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 404  
|||||

RESULT 11  
AZ994818/c

LOCUS 721 bp DNA linear GSS 27-APR-2001  
DEFINITION clone UUGC2M0280J24 F, genomic survey sequence.

ACCESSION AZ994818  
VERSION 1  
KEYWORDS GSS.

SOURCE AZ994818.1 GI:13866045  
Mus musculus (house mouse)

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 721)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,  
and Wright, D., Weis, R.

## TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
COMMENT

plasmid inserts  
Unpublished

Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0280 row: J column: 24  
Seq primer: CGTTGTAACACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 721.

## FEATURES

source

Location/Qualifiers  
1..721  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0280J24"  
/sex="female"  
/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/notes="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (GI4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 191 a 202 c 166 g 162 t  
ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 721;  
Best Local Similarity 100.0%; Pred. No. 1.8e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCTCTGTGGAAGG 60  
|||||

DB 300 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCTCTGTGGAAGG 241  
|||||

QY 61 CAGCTCTGATGCATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAAACGACAG 120  
|||||

DB 240 CAGCTCTGATGCATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAAACGACAG 181  
|||||

QY 121 GGAGCTGGCGACAGACCTCTCTGGGTGATGAGCTTAAGGATGTTTGTGTAGGGCCCT 180  
|||||

DB 180 GGAGCTGGCGACAGACCTCTCTGGGTGATGAGCTTAAGGATGTTTGTGTAGGGCCCT 121  
|||||

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
|||||

DB 120 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 70  
|||||

RESULT 12

AZ310386/c

LOCUS 472 bp DNA linear GSS 29-SEP-2000  
DEFINITION clone UUGC1M0025013 F, genomic survey sequence.

A2310386  
 VERSION A2310386.1 GI:10352324  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0025 row: O column: 13  
 Seq primer: CGGTGTAAACGACGGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 472.

FEATURES  
 Source  
 1..472  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UGCCIM0025013"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UGCLM library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydronically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (gii4732114(gb)AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 133 a 140 c 104 g 95 t  
 ORIGIN  
 Query Match 99.3%; Score 229.4; DB 28; Length 472;  
 Best Local Similarity 99.6%; Pred. No. 4.6e-60;  
 Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTGTGGAAGG 60  
 Db 281 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTGTGGAAGG 222  
 QY 61 CAGCTCTGATTCATGAAGGTTTCAGTCTCCTAGTTCCTCCCGAGGAAAAACGACACG 120  
 Db 221 CATGCTGATTCATGAAGGTTTCAGTCTCCTAGTTCCTCCCGAGGAAAAACGACACG 162  
 QY 121 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTAGGGCCCT 180

Db 161 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTAGGGCCCT 102  
 QY 181 ATGCTTGCACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231  
 Db 101 ATGCTTGCACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 51  
 RESULT 13  
 LOCUS AZ054649 505 bp DNA linear GSS 30-MAR-2000  
 DEFINITION RPCI-23-401C1.TJ RPCI-23 Mus musculus genomic clone RPCI-23-401C1, genomic survey sequence.  
 ACCESSION AZ054649  
 VERSION AZ054649.1 GI:7345885  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akineret,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P., and Fraser,C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL Unpublished  
 COMMENT Other GSSs: RPCI-23-401C1.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 401 row: C column: 1  
 Seq primer: SP6  
 Class: BAC ends.  
 Location/Qualifiers  
 source  
 1..505  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-401C1"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /clone\_lib="RPCI-23"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 100 a 114 c 147 g 144 t  
 ORIGIN  
 Query Match 99.3%; Score 229.4; DB 28; Length 505;  
 Best Local Similarity 99.6%; Pred. No. 4.8e-60;  
 Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTGTGGAAGG 60  
 Db 233 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTGTGGAAGG 292  
 QY 61 CAGCTCTGATTCATGAAGGTTTCAGTCTCCTAGTTCCTCCCGAGGAAAAACGACACG 120



```

Db      293 CATGCTGATGTCAGAGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACGACACG 352
QY      121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGATGTTTGTGTAGGCCCT 180
        |||||||
Db      353 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGATGTTTGTGTAGGCCCT 412
QY      181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
        |||||||
Db      413 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 463

RESULT 14
AW558986/c
LOCUS   AW558986
DEFINITION L0302F10-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
clone L0302F10 3', mRNA sequence.
ACCESSION AW558986
VERSION   AW558986
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 518)
AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
TITLE     Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE   20381348
PUBMED    10922068
COMMENT   Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: L0302 row: F column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 518
POLYA=Yes.

FEATURES
         Location/Qualifiers
            1..518
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="niaEST:L0302F10-3"
                /db_xref="taxon:10090"
                /clone="L0302F10"
                /sex="female"
                /dev_stage="Newborn Ovary"
                /lab_host="DH10B"
                /clone_lib="NIA Mouse Newborn Ovary cDNA Library"
                /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
                Site.1: Salt; Site.2: NotI; Total RNAs were extracted from
                7 Newborn Ovary. The double-stranded cDNA was synthesized
                by Gibco's kit with an oligo(dT) primer [NotI
                primer-adaptor from GibcoBRL]
                [5'-pGACAGTCTAGATCGGCGCGCCGCTTTTCTTTT-3'] from
                2.56ug of total RNA. The double-stranded cDNAs were
                treated with T4 DNA polymerase and purified by
                ethanol-precipitation. The cDNAs were ligated to
                lone-linker LL-Sal3 (include SalI sequence). The cDNAs
                were purified by phenol/chloroform and separated from
                free linkers by Centricon 100. Then, cDNAs were amplified
                by long-range high fidelity PCR using Takara's Ex Taq
                polymerase. Then, the cDNAs were purified by
                phenol/chloroform and by Centricon 100. The cDNAs were
                digested with SalI and NotI enzymes. Then, the cDNAs were
                size selected by Gibco's Size Fractionation Column. The
                cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
                vector. The DH10B E. coli host was transformed with the
                ligation mixture by chemical method. The library was

```

```

BASE COUNT 126 a 138 c 131 g 123 t
ORIGIN
Query Match 99.3%; Score 229.4; DB 9; Length 518;
Best Local Similarity 99.6%; Pred. No. 4.9e-60;
Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GTGGGTGCGAGGCTAAGCAGCTACGACGAGGATAGCTTGTGTGCATCCTGTGGAAGG 60
        |||||||
Db      465 GTGGGTGCGAGGCTAAGCAGCTACGACGAGGATAGCTTGTGTGCATCCTGTGGAAGG 406
QY      61 CAGCTCTGATTCGATGATGAAGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACGACACG 120
        |||||||
Db      405 CAGCTCTGATTCGATGATGAAGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACGACACG 346
QY      121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGATGTTTGTGTAGGCCCT 180
        |||||||
Db      345 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGATGTTTGTGTAGGCCCT 286
QY      181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
        |||||||
Db      285 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCGCCCCATGAGGCTTGCTT 235

RESULT 15
BM219777/c
LOCUS   BM219777
DEFINITION C0929H06-3 NIA Mouse 12.5-dpc Male Genital Ridge/Mesonephros cDNA
Library (Long) Mus musculus cDNA clone NIA:C0929H06 IMAGE:30037433
3', mRNA sequence.
ACCESSION BM219777
VERSION   BM219777.2
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 520)
AUTHORS   Piao,Y., KO,N.T., Lim,M.K. and Ko,M.S.H.
TITLE     Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE   21429098
PUBMED    11544199
COMMENT   On Dec 14, 2001 this sequence version replaced gi:17779702.
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: C0929 row: H column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 520
POLYA=Yes.

FEATURES
         Location/Qualifiers
            1..520
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="niaEST:C0929H06-3"
                /db_xref="taxon:10090"
                /clone="NIA:C0929H06 IMAGE:30037433"
                /sex="Male"
                /tissue_type="Male genital ridge/mesonephros"
                /dev_stage="12.5-dpc"
                /lab_host="DH10B"
                /clone_lib="NIA Mouse 12.5-dpc Male Genital
                Ridge/Mesonephros cDNA Library (Long)"
                /note="Vector: pSPORT1 (Invitrogen); Site.1: Salt; Site.2:
                NotI; Mouse cDNA project by the Laboratory of Genetics,
                National Institute on Aging (NIA), Intramural Research

```

Program, NIH (<http://lgsun.grc.nia.nih.gov/cdna>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199]. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]:  
5'-TGACTAGTCTAGATCGGAGCGCGCCCTTTTTTTTTTTT-3' from 1.8 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 122 a 140 c 134 g 124 t  
ORIGIN

Query Match 99.3%; Score 229.4; DB 12; Length 520;  
Best Local Similarity 99.6%; Pred. No. 4.9e-60;  
Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAAGCAGCTGACAGAGGATAGCTTGGTGGCATCTCTGTGAAGG 60  
Db 465 GTGGGTGCGAGGCTAAGCAGCTGACAGAGGATAGCTTGGTGGCATCTCTGTGAAGG 406

Qy 61 CAGCTCTGATGCGATGAGGCTTCAAGTCTAGTTCCTCCCTCCCGAGGAAAAAGACAG 120  
Db 405 CAGCTCTGATGCGATGAGGCTTCAAGTCTAGTTCCTCCCTCCCGAGGAAAAAGACAG 346

Qy 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTGTAGGGCCCT 180  
Db 345 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTGTAGGGCCCT 286

Qy 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
Db 285 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 235

Search completed: October 9, 2003, 22:40:58  
Job time : 2067 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:44:04 ; Search time 206 Seconds  
(without alignments)  
2909.716 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtaggggtaggagctaaagca.....ttcaccatgaggtgtgtt 231

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09A\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
C 1	214.2	92.7	185548	14	US-10-175-523-62	Sequence 62, Appl
C 2	92.6	40.1	432	10	US-09-933-797-502	Sequence 502, App
C 3	89.4	38.7	1913	12	US-10-225-810-40	Sequence 40, Appl
C 4	86.2	37.3	1434	10	US-09-799-462A-13	Sequence 13, Appl
C 5	86.2	37.3	1434	11	US-09-836-911A-13	Sequence 13, Appl
C 6	86.2	37.3	1434	13	US-10-125-767-13	Sequence 13, Appl
C 7	86.2	37.3	1434	14	US-10-151-081-13	Sequence 13, Appl
C 8	86.2	37.3	1434	14	US-10-287-313-13	Sequence 13, Appl
C 9	86.2	37.3	1434	14	US-10-219-694-13	Sequence 13, Appl
C 10	81.4	35.2	1400	11	US-09-799-462A-14	Sequence 14, Appl
C 11	81.4	35.2	1400	11	US-09-836-911A-14	Sequence 14, Appl
C 12	81.4	35.2	1400	13	US-10-125-767-14	Sequence 14, Appl
C 13	81.4	35.2	1400	14	US-10-151-081-14	Sequence 14, Appl
C 14	81.4	35.2	1400	14	US-10-287-313-14	Sequence 14, Appl
C 15	81.4	35.2	1400	14	US-10-219-694-14	Sequence 14, Appl
C 16	70.4	30.5	251364	14	US-10-175-523-62	Sequence 58, Appl

Sequence 61, Appl	US-10-175-523-61	14	251364	30.5	70.4	17
Sequence 76, Appl	US-10-175-523-79	14	251364	30.5	70.4	18
Sequence 39, Appl	US-09-563-728A-36	11	14.7	34	34	19
Sequence 4575, Ap	US-09-867-701-4575	10	475	13.2	30.4	20
Sequence 39036, A	US-10-027-632-39036	13	447	13.0	30	21
Sequence 65139, A	US-10-027-632-65139	13	447	13.0	30	22
Sequence 2639, Ap	US-09-918-995-2639	11	457	13.0	30	23
Sequence 112, App	US-10-231-417-112	12	609	13.0	30	24
Sequence 151950,	US-10-027-632-151950	13	968	12.8	29.6	25
Sequence 151951,	US-10-027-632-151951	13	968	12.8	29.6	26
Sequence 151952,	US-10-027-632-151952	13	968	12.8	29.6	27
Sequence 181112,	US-10-027-632-181112	13	499	12.7	29.4	28
Sequence 109, App	US-10-102-627-109	14	32187	12.7	29.4	29
Sequence 95, Appl	US-09-250-611-95	10	540	12.6	29.2	30
Sequence 112118,	US-10-027-632-112118	13	2977	12.6	29.2	31
Sequence 112119,	US-10-027-632-112119	13	2977	12.6	29.2	32
Sequence 10697, A	US-09-783-590-10697	10	360	12.4	28.6	33
Sequence 1096, Ap	US-09-764-847-1096	14	427	12.4	28.6	34
Sequence 1096, Ap	US-10-092-154-1096	14	427	12.4	28.6	35
Sequence 1086, Ap	US-10-027-632-70833	13	564	12.4	28.6	36
Sequence 70833, A	US-10-027-632-70834	13	564	12.4	28.6	37
Sequence 70834, A	US-10-027-632-70834	13	564	12.4	28.6	38
Sequence 312268,	US-10-027-632-312268	13	564	12.4	28.6	39
Sequence 312269,	US-10-027-632-312269	13	564	12.4	28.6	40
Sequence 5, Appl	US-10-080-381B-5	12	1197	12.4	28.6	41
Sequence 45, Appl	US-10-240-985-45	12	6549	12.4	28.6	42
Sequence 102751,	US-10-027-632-102751	13	2449	12.3	28.4	43
Sequence 112101,	US-10-027-632-112101	13	2449	12.3	28.4	44
Sequence 1615, Ap	US-09-764-847-1615	10	10472	12.3	28.4	45
Sequence 1615, Ap	US-10-092-154-1615	14	10472	12.3	28.4	46

## ALIGNMENTS

RESULT 1  
US-10-175-523-62/c  
; Sequence 62, Application US/10175523  
; Publication No. US20030096264A1  
; GENERAL INFORMATION:  
; APPLICANT: Brockman, Jeffrey  
; APPLICANT: Evans, David  
; APPLICANT: Hook, Derek  
; APPLICANT: Klimczak, Leszek  
; APPLICANT: Laeng, Pascal  
; APPLICANT: Palfreyman, Michael  
; APPLICANT: Rajan, Priithi  
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)  
; FILE REFERENCE: 3235/1J795-US3  
; CURRENT APPLICATION NUMBER: US/10/175,523  
; PRIOR FILING DATE: 2002-06-18  
; PRIOR APPLICATION NUMBER: US 60/299,151  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: US 60/317,828  
; PRIOR FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: US 60/325,150  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/333,047  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: US 60/349,936  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US 60/361,834  
; PRIOR FILING DATE: 2002-03-04  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62  
; LENGTH: 185548  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-175-523-62

Query Match 92.7%; Score 214.2; DB 14; Length 185548;  
Best Local Similarity 98.3%; Pred. No. 1.3e-67;  
Matches 227; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 GTGGGTCGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCCTGTGGGAAG 60  
Db 52312 GTGGGTCGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCCTGTGGGAAG 52253  
QY 61 CACGCTCGATTGCAATGAAGTTCAGTGTCTAGTTCCTTCCGCCAGAAAAAGACAGC 120  
Db 52252 TAGTCTGCA-TCCATGAAGGTTCAGTGTCTAGTTCCTTCCGCCAGAAAAAGACAGC 52194  
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTGTGTAGGGCCCT 180  
Db 52193 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTGTGTAGGGCCCT 52134  
QY 181 ATGCTTCACACATGGGGATCAGACCTCTACCTTACCCCATGAGCTTGTCTT 231  
Db 52133 ATGCTTCACACATGGGGATCAGACCTCTACCTTACCCCATGAGCTTGTCTT 52083

## RESULT 2

US-09-933-797-502/c  
; Sequence 502, Application US/09933797  
; Patent No. US20020155119A1  
; GENERAL INFORMATION:  
; APPLICANT: Robert A. Sikes et al.  
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital  
; TITLE OF INVENTION: Sinus Expressed Sequences  
; FILE REFERENCE: 9901-007-999  
; CURRENT APPLICATION NUMBER: US/09/933,797  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: US/09/482,933  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: PCT/US99/10746  
; PRIOR FILING DATE: 1999-05/14  
; PRIOR APPLICATION NUMBER: 60/085,383  
; PRIOR FILING DATE: 1998-05-14  
; NUMBER OF SEQ ID NOS: 811  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 502  
; LENGTH: 432  
; TYPE: DNA  
; ORGANISM: Murine  
US-09-933-797-502

Query Match 40.1%; Score 92.6; DB 10; Length 432;  
Best Local Similarity 76.0%; Pred. No. 9.9e-24;  
Matches 130; Conservative 0; Mismatches 34; Indels 7; Gaps 1;  
QY 55 GGAAGGCACGCTCTGATTGCAATGAAGTTCAGTGTCTAGTTCCTTCCGCCAGAAAAAC 114  
Db 416 GGGAGACATGTCATCTTTCATGAGGTTCAGTGTCTAGTTCCTTCCGCCAGAAAAAC 357  
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTTAAGGGATGTTT 167  
Db 356 GACACGGGAGCAGTCAAGGTTGCTCTGGGTAAAAGCCTGAGCCTTAAGAGCTAATCCT 297  
QY 168 GTGTAGGGCCCTATGCTTCACACATGGGGATCAGACCTTACCTTACCC 218  
Db 296 GTACATGGCTCTTACCTACACACTGGGGATTGACCTTATCTCCACTC 246

## RESULT 3

US-10-225-810-40/c  
; Sequence 40, Application US/10225810  
; Publication No. US20030157512A1  
; GENERAL INFORMATION:  
; APPLICANT: Bermingham, Jr., John R.  
; TITLE OF INVENTION: Transdorins and Methods of Using Transdorin  
; FILE REFERENCE: McLaugh-07165  
; CURRENT APPLICATION NUMBER: US/10/225,810  
; CURRENT FILING DATE: 2002-08-21  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 40

; LENGTH: 1913  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-225-810-40

Query Match 38.7%; Score 89.4; DB 12; Length 1913;  
Best Local Similarity 74.9%; Pred. No. 2.2e-22;  
Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;  
QY 55 GGAAGGCACGCTCTGATTGCAATGAAGTTCAGTGTCTAGTTCCTTCCGCCAGAAAAAC 114  
Db 1347 GGGAGACATGTCATCTTTCATGAGGTTCAGTGTCTAGTTCCTTCCGCCAGAAAAAC 1288  
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTTAAGGGATGTTT 167  
Db 1287 GACACGGGAGCAGTCAAGGTTGCTCTGGGTAAAAGCCTTAAGCCTTAAGAGCTAATCCT 1228  
QY 168 GTGTAGGGCCCTATGCTTCACACATGGGGATCAGACCTTACCTTACCC 218  
Db 1227 GTACATGGCTCTTACCTACACACTGGGGATTGACCTTATCTCCACTC 1177

## RESULT 4

US-09-799-462A-13  
; Sequence 13, Application US/09799462A  
; Patent No. US20020160970A1  
; GENERAL INFORMATION:  
; APPLICANT: Hadlaczky, Gyula  
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF  
; AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/799,462A  
; FILING DATE: 10-Sep-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/835,682  
; FILING DATE: 10-APR-1997  
; APPLICATION NUMBER: 08/695,191  
; FILING DATE: 07-AUG-1996  
; APPLICATION NUMBER: 08/682,080  
; FILING DATE: 15-JUL-1996  
; APPLICATION NUMBER: 08/629,822  
; FILING DATE: 10-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24601-402G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 858-450-8403  
; TELEFAX: 858-587-5360  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1434 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO

ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-799-462A-13

Query Match 37.3%; Score 86.2; DB 10; Length 1434;  
Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;  
QY 55 GGAAGGCACGCTGATGTCATGAAGGTCAGTGTCTAGTTCCTCCCTCCCGAGGAAAAC 114  
DB 900 GGGGACATGTCATCTTTCATGAAGGTCAGTGTCTAGTTCCTCCCTCCCGAGGAAAAC 959  
QY 115 GACAGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGATGTTT 167  
DB 960 GACAGGGAGCTGGCCAGGTCAGGTTGCTCTGGTAAAGCCCTGAGCCTGGAGCTAATCT 1019  
QY 168 GGTAGGGCCCTATGCTTGCACACTGGGGATTCACCTCTACCTTCACCC 218  
DB 1020 GTACATGGCTCCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 1070

## RESULT 5

US-09-836-911A-13  
Sequence 13, Application US/09836911A  
Publication No. US20030033617A1  
GENERAL INFORMATION:

APPLICANT: Hadlaczky, Gyula

Szalay, Aladar

TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe

STREET: 4350 La Jolla Village Drive, 6th Floor

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/836,911A

FILING DATE: 17-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835,682

FILING DATE: 10-Apr-1997

APPLICATION NUMBER: 08/695,191

FILING DATE: 07-Aug-1996

APPLICATION NUMBER: 08/682,080

FILING DATE: 15-Jul-1996

APPLICATION NUMBER: 08/629,822

FILING DATE: 10-Apr-1996

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24601-4021

TELECOMMUNICATION INFORMATION:

TELEPHONE: 858-450-8403

TELEFAX: 858-587-5360

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 1434 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-836-911A-13

Query Match 37.3%; Score 86.2; DB 11; Length 1434;  
Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;  
QY 55 GGAAGGCACGCTGATGTCATGAAGGTCAGTGTCTAGTTCCTCCCTCCCGAGGAAAAC 114  
DB 900 GGGGACATGTCATCTTTCATGAAGGTCAGTGTCTAGTTCCTCCCTCCCGAGGAAAAC 959  
QY 115 GACAGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGATGTTT 167  
DB 960 GACAGGGAGCTGGCCAGGTCAGGTTGCTCTGGTAAAGCCCTGAGCCTGGAGCTAATCT 1019  
QY 168 GGTAGGGCCCTATGCTTGCACACTGGGGATTCACCTCTACCTTCACCC 218  
DB 1020 GTACATGGCTCCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 1070

## RESULT 6

US-10-125-767-13  
Sequence 13, Application US/10125767  
Publication No. US20020160410A1  
GENERAL INFORMATION:

APPLICANT: Hadlaczky, Gyula

Szalay, Aladar

TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe LLP

STREET: 4350 La Jolla Village Drive, 7th Floor

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/125,767

FILING DATE: 17-Apr-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/724,693

FILING DATE: 28-Nov-2000

APPLICATION NUMBER: 08/835,682

FILING DATE: 10-Apr-1997

APPLICATION NUMBER: 08/695,191

FILING DATE: 07-Aug-1996

APPLICATION NUMBER: 08/682,080

FILING DATE: 15-Jul-1996

APPLICATION NUMBER: 08/629,822

FILING DATE: 10-Apr-1996

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24601-4021

TELECOMMUNICATION INFORMATION:

TELEPHONE: 858-450-8403

TELEFAX: 858-587-5360

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 1434 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: Genomic DNA



; TELEPHONE: 858-450-8403  
; TELEFAX: 858-587-5360  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1434 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: <Unknown>  
; ORIGINAL SOURCE:  
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-10-287-313-13

Query Match 37.3%; Score 86.2; DB 14; Length 1434;  
Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;  
QY 55 GGAAGCAGCTCTGATTCGATGAAGTTTCAGTCTCAGTCTCCCTTCCCGAGGAAAAAC 114  
DB 900 GGGGACATGTCATCTTTCATGAAGTTTCAGTCTCAGTCTCCCTTCCCGAGGAAAAAC 959  
QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTTT 167  
DB 960 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019  
QY 168 GTGTAGGGCCCTTATCTTGCACACATGGGGATCAGACCTCTACCTTCACCC 218  
DB 1020 GTACATGGCTCTTTACCTACACATGGGGATTGACCTCTATCTCCACTC 1070

## RESULT 9

US-10-219-694-13  
; Sequence 13, Application US/10219694  
; Publication No. US20030108914A1  
; GENERAL INFORMATION:  
; APPLICANT: Hadlaczky, Gyula  
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS  
; FOR PREPARING ARTIFICIAL CHROMOSOMES  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
; STREET: 4350 La Jolla Village Drive, 7th Floor  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/219,694  
; FILING DATE: 14-Aug-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: 28-NOV-2000  
; APPLICATION NUMBER: 08/835,682  
; FILING DATE: 10-APR-1997  
; APPLICATION NUMBER: 08/695,191  
; FILING DATE: 07-AUG-1996  
; APPLICATION NUMBER: 08/682,080  
; FILING DATE: 15-JUL-1996  
; APPLICATION NUMBER: 08/629,822  
; FILING DATE: 10-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779

; REFERENCE/DOCKET NUMBER: 24601-402M  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 858-450-8403  
; TELEFAX: 858-587-5360  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1434 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: <Unknown>  
; ORIGINAL SOURCE:  
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-10-219-694-13

Query Match 37.3%; Score 86.2; DB 14; Length 1434;  
Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;  
QY 55 GGAAGCAGCTCTGATTCGATGAAGTTTCAGTCTCAGTCTCCCTTCCCGAGGAAAAAC 114  
DB 900 GGGGACATGTCATCTTTCATGAAGTTTCAGTCTCAGTCTCCCTTCCCGAGGAAAAAC 959  
QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTTT 167  
DB 960 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019  
QY 168 GTGTAGGGCCCTTATCTTGCACACATGGGGATCAGACCTCTACCTTCACCC 218  
DB 1020 GTACATGGCTCTTTACCTACACATGGGGATTGACCTCTATCTCCACTC 1070

## RESULT 10

US-09-799-462A-14  
; Sequence 14, Application US/09799462A  
; Patent No. US20020160970A1  
; GENERAL INFORMATION:  
; APPLICANT: Hadlaczky, Gyula  
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF  
; AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/799,462A  
; FILING DATE: 10-Sep-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/835,682  
; FILING DATE: 10-APR-1997  
; APPLICATION NUMBER: 08/695,191  
; FILING DATE: 07-AUG-1996  
; APPLICATION NUMBER: 08/682,080  
; FILING DATE: 15-JUL-1996  
; APPLICATION NUMBER: 08/629,822  
; FILING DATE: 10-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L

```
;
;
;   REGISTRATION NUMBER: 33,779
;   REFERENCE/DOCKET NUMBER: 24601-402G
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 858-450-8403
;   TELEFAX: 858-587-5360
;   TELEX: <Unknown>
;
;   INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1400 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: <Unknown>
;   ORIGINAL SOURCE:
;   SEQUENCE DESCRIPTION: SEQ ID NO: 14:
;
; US-09-799-462A-14
;
; Query Match          35.2%   Score 81.4; DB 10; Length 1400;
; Best Local Similarity 71.9%; Pred. No. 1.7e-19;
; Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
;
; QY      55 GGAAGGCACCTCTGATTGCATGAAGTTCAGTGTCTAGTTCCTTCCCTCCCGGAAAAAC 114
;         || || || || || || || || || || || || || || || || || || || || ||
; Db      771 GGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCCAAAGTGTCTTCTCCCGGAAAAAC 830
;
; QY      115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAAGGGATGTTT 167
;         || || || || || || || || || || || || || || || || || || || || ||
; Db      831 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTGTGAGCCCTAAAGCTAATCCT 890
;
; QY      168 GTGTAGGCCCCCTATGCTTGCACACTGGGATCAGACCTCTACCTTCCACCC 218
;         || || || || || || || || || || || || || || || || || || || || ||
; Db      891 GTACATGGCTCCCTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 941
;
;
; RESULT 11
; US-09-836-911A-14
;   Sequence 14, Application US/09836911A
;   Publication No. US20030033617A1
;   GENERAL INFORMATION:
;   APPLICANT: Hadlaczky, Gyula
;   TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;   AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;
;   NUMBER OF SEQUENCES: 34
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Heller Ehrman White & McAuliffe
;   STREET: 4350 La Jolla Village Drive, 6th Floor
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92122
;
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSeq Version 1.5
;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/836,911A
;   FILING DATE: 17-Apr-2002
;   CLASSIFICATION: <Unknown>
;
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/835,682
;   FILING DATE: 10-APR-1997
;   APPLICATION NUMBER: 08/695,191
;   FILING DATE: 07-AUG-1996
;   APPLICATION NUMBER: 08/682,080
;   FILING DATE: 15-JUL-1996
;   APPLICATION NUMBER: 08/629,822
;   FILING DATE: 10-APR-1996
;
;   ATTORNEY/AGENT INFORMATION:
```

```
;
;
;   NAME: Seidman, Stephanie L
;   REGISTRATION NUMBER: 33,779
;   REFERENCE/DOCKET NUMBER: 24601-402I
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 858-450-8403
;   TELEFAX: 858-587-5360
;   TELEX: <Unknown>
;
;   INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1400 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: <Unknown>
;   ORIGINAL SOURCE:
;   SEQUENCE DESCRIPTION: SEQ ID NO: 14:
;
; US-09-836-911A-14
;
; Query Match          35.2%   Score 81.4; DB 11; Length 1400;
; Best Local Similarity 71.9%; Pred. No. 1.7e-19;
; Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
;
; QY      55 GGAAGGCACCTCTGATTGCATGAAGTTCAGTGTCTAGTTCCTTCCCTCCCGGAAAAAC 114
;         || || || || || || || || || || || || || || || || || || || || ||
; Db      771 GGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCCAAAGTGTCTTCTCCCGGAAAAAC 830
;
; QY      115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAAGGGATGTTT 167
;         || || || || || || || || || || || || || || || || || || || || ||
; Db      831 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTGTGAGCCCTAAAGCTAATCCT 890
;
; QY      168 GTGTAGGCCCCCTATGCTTGCACACTGGGATCAGACCTCTACCTTCCACCC 218
;         || || || || || || || || || || || || || || || || || || || || ||
; Db      891 GTACATGGCTCCCTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 941
;
;
; RESULT 12
; US-10-125-767-14
;   Sequence 14, Application US/10125767
;   Publication No. US20020160410A1
;   GENERAL INFORMATION:
;   APPLICANT: Hadlaczky, Gyula
;   TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
;   METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;
;   NUMBER OF SEQUENCES: 34
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Heller Ehrman White & McAuliffe LLP
;   STREET: 4350 La Jolla Village Drive, 7th Floor
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92122
;
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSeq Version 1.5
;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/125,767
;   FILING DATE: 17-Apr-2002
;
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/724,693
;   FILING DATE: 28-NOV-2000
;   APPLICATION NUMBER: 08/835,682
;   FILING DATE: 10-APR-1997
;   APPLICATION NUMBER: 08/695,191
;   FILING DATE: 07-AUG-1996
;   APPLICATION NUMBER: 08/682,080
;   FILING DATE: 15-JUL-1996
;
;   APPLICATION NUMBER: 08/629,822
```



```

; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-151-081-14

Query Match 35.2%; Score 81.4; DB 14; Length 1400;
Best Local Similarity 71.9%; Pred. No. 1.7e-19;
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

QY 55 GGAAGGCACCTCGATTGCATGAAGGTTCAGTCTCCTAGTTCCTCCCTCCCGAGAAAC 114
DB 771 GGGAGACATGTCATCTTTCAAGAAGGTTCAGTCTCCTCCCTCCCGAGAAAC 830

QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTTT 167
DB 831 GACACGGGAGCAGCTCAGGGTTCTCTGGGTAAAGCCCTGTGAGCCTAAGAGCTAATCCT 890

QY 168 GTCTAGGGCCCTATGCTTGCACACACATGGGGATCAGACCTCTACCTTCACCC 218
DB 891 GTACATGGCTCTTTACCTACACACATGGGGATTTGACCTCTATCTCCACTC 941

RESULT 14
US-10-287-313-14
; Sequence 14, Application US/10287313
; Publication No. US20030101480A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & Mcauliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/287,313
; FILING DATE: 01-No. US20030101480A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/724,726

```

```

; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6869-402N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-287-313-14

```

```

Query Match 35.2%; Score 81.4; DB 14; Length 1400;
Best Local Similarity 71.9%; Pred. No. 1.7e-19;
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

QY 55 GGAAGCAGCTGTGATTGATGAAGGTTTCAGTCTCCTTCCCTCCCGAGGAAAAAC 114
Db 771 GGGAGACATGTCATCTTCAAGAAGGTTGAGTGTCAGAGTGCTTCTCCAGGCAAAAC 830
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTT 167
Db 831 GACACGGGAGCAGGTGCTTCAAGAAGGTTGAGTGTCAGAGTGCTTCTCCAGGCAAAAC 830
QY 168 GTGAGGGCCCTATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218
Db 891 GTACATGGCTCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 941

```

```

RESULT 15
US-10-219-694-14
; Sequence 14, Application US/10219694
; Publication No. US20030108914A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
; FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP
; STREET: 4350 La Jolla Village Drive, 7th Floor
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/219,694
; FILING DATE: 14-Aug-2002
; CLASSIFICATION: <Unknown>

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402M
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-219-694-14

```

```

Query Match 35.2%; Score 81.4; DB 14; Length 1400;
Best Local Similarity 71.9%; Pred. No. 1.7e-19;
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

QY 55 GGAAGCAGCTGTGATTGATGAAGGTTTCAGTCTCCTTCCCTCCCGAGGAAAAAC 114
Db 771 GGGAGACATGTCATCTTCAAGAAGGTTGAGTGTCAGAGTGCTTCTCCAGGCAAAAC 830
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTT 167
Db 831 GACACGGGAGCAGGTGCTTCAAGAAGGTTGAGTGTCAGAGTGCTTCTCCAGGCAAAAC 830
QY 168 GTGAGGGCCCTATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218
Db 891 GTACATGGCTCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 941

```

Search completed: October 9, 2003, 23:40:57  
Job time : 208 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:06:28 ; Search time 73 Seconds

(without alignments)  
1396.706 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 ggggggtgcaggctaaagca.....ttcaccatgaggtgtgctt 231

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2.6/ptodata/1/lna/5A\_COMB.seq.\*

2: /cgn2.6/ptodata/1/lna/5B\_COMB.seq.\*

3: /cgn2.6/ptodata/1/lna/6A\_COMB.seq.\*

4: /cgn2.6/ptodata/1/lna/6B\_COMB.seq.\*

5: /cgn2.6/ptodata/1/lna/PTUS\_COMB.seq.\*

6: /cgn2.6/ptodata/1/lna/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.2	37.3	1434	3	US-08-695-191-13
2	86.2	37.3	1434	3	US-08-682-080-13
3	81.4	35.2	1400	3	US-08-695-191-14
4	81.4	35.2	1400	3	US-08-682-080-14
5	35	15.2	50	4	US-08-585-593A-4
6	29.2	12.6	540	4	US-09-250-609-95
7	29.2	12.6	540	4	US-09-250-611-95
8	28.6	12.4	2692	1	US-07-932-454A-2
9	28	12.1	866	4	US-09-620-312D-682
10	27.6	11.9	1865	3	US-08-766-354A-29
11	27.2	11.8	2297	2	US-08-588-983-21
12	27.2	11.8	2297	2	US-08-588-976-21
13	27	11.7	1054	6	5189147-4
14	26.8	11.6	503	4	US-09-250-609-85
15	26.8	11.6	503	4	US-09-250-611-85
16	26.8	11.6	3213	2	US-08-633-770A-4
17	26.4	11.4	13865	3	US-09-009-217-11
18	26.4	11.4	13865	3	US-09-009-656-11
19	26.2	11.3	71	1	US-08-434-001-210
20	26.2	11.3	71	1	US-08-433-585-210
21	26.2	11.3	71	1	US-08-434-425-210
22	26.2	11.3	71	2	US-08-437-667-210
23	26.2	11.3	71	3	US-08-906-955-210
24	26.2	11.3	71	3	US-08-945-909-210
25	26.2	11.3	71	4	US-09-396-002A-210
26	26.2	11.3	71	5	PT-US96-06060-210
27	26	11.3	66804	4	US-09-740-041-3

28 11.3 536165 4 US-09-214-808-1 Sequence 1, Appli  
 29 25.8 444 3 US-08-928-213B-180 Sequence 180, App  
 30 25.8 866 4 US-09-495-050A-55 Sequence 55, Appl  
 31 25.8 3883 4 US-09-620-312D-792 Sequence 792, App  
 32 25.8 5865 3 US-09-011-745-8 Sequence 8, Appli  
 33 25.8 9058 3 US-08-913-014A-9 Sequence 9, Appli  
 34 25.8 1830121 4 US-09-557-884-1 Sequence 1, Appli  
 35 25.8 1830121 4 US-09-643-990A-1 Sequence 1, Appli  
 36 25.6 11.1 419 3 US-09-214-095D-113 Sequence 113, App  
 37 25.6 11.1 504 4 US-09-328-475C-188 Sequence 188, App  
 38 25.6 11.1 4168 4 US-09-266-225D-17 Sequence 17, Appl  
 39 25.6 11.1 4279 3 US-09-041-886-22 Sequence 22, Appl  
 40 25.6 11.1 319608 4 US-09-539-333D-1 Sequence 1, Appli  
 41 25.6 11.1 319608 4 US-09-679-409-1 Sequence 1, Appli  
 42 25.4 11.0 4643 4 US-09-436-874-1 Sequence 1, Appli  
 43 25.4 11.0 4403765 3 US-09-103-840A-2 Sequence 2, Appli  
 44 25.4 11.0 4411529 3 US-09-103-840A-1 Sequence 1, Appli  
 45 25.2 10.9 371 1 US-08-620-467A-1 Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
 US-08-695-191-13  
 ; Sequence 13, Application US/08695191  
 ; Patent No. 6025155  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hadlaczky, Gyula  
 ; APPLICANT: Szalay, Aladar  
 ; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND  
 ; METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Brown, Martin, Haller & McClain  
 ; STREET: 1660 Union Street  
 ; CITY: San Diego  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92101-2926  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/695,191  
 ; FILING DATE: 07-AUG-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE: 15-JUL-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA: 08/629,822  
 ; APPLICATION NUMBER:  
 ; FILING DATE: 10-APR-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seidman, Stephanie L.  
 ; REGISTRATION NUMBER: 33,779  
 ; REFERENCE/DOCKET NUMBER: 6869-402C  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-238-0999  
 ; TELEFAX: 619-238-0062  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 13:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1434 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; HYPOTHETICAL: NO

```

: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
US-08-695-191-13

Query Match 37.3% Score 86.2; DB 3; Length 1434;
Best Local Similarity 73.7%; Pred. No. 1.3e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

QY 55 GGAAGGCACGCTCGATTGCATGAAGGTTTCAGTGTCTTCCCTTCCCTCCAGGAAAAAC 114
Db 900 GGGGACATGTCATCTTTTCATGAAGGTTTCAGTGTCTTCCCTTCCCTCCAGGAAAAAC 959
QY 115 GACACGGGAGCTGGCCAAAGACCTCTCTGGGTGA-----TGAGCCTTAAGGGATGGTTTT 167
Db 960 GACACGGGAGCAGGTCAGGGTTGCTTGGGTAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019
QY 168 GTGTAGGCGCCCTTCGTTCCACACTGGGATCAGACCTCTACTTTCACCC 218
Db 1020 GTACATGGCTCCTTTACCTACACACTGGGATTTGACCTCTATCTCACATC 1070

RESULT 2
US-08-682-080-13
: Sequence 13, Application US/08682080
: Patent No. 6077697
: GENERAL INFORMATION:
: APPLICANT: Hadlaczky, Gyula
: APPLICANT: Szalay, Aladar
: TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR
: TITLE OF INVENTION: PREPARING ARTIFICIAL CHROMOSOMES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92101-2926
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,080
: FILING DATE: 10-APR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/629,822
: FILING DATE: 10-APR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Seidman, Stephanie L
: REGISTRATION NUMBER: 33,779
: REFERENCE/DOCKET NUMBER: 6869-402B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-238-0999
: TELEFAX: 619-238-0062
: TELEX:
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1434 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
US-08-682-080-13

```

```

Query Match          37.3%; Score 86.2; DB 3; Length 1434;
Best Local Similarity 73.7%; Pred. No. 1.3e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

QY      55   GGAGGGCAGCTCTGATTGCATGAAGGTTCAGTGTCCCTAGTTCCCTTCCCCCAGGAAAC 114
         || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      900   GGGGGAGCATGTCATCTTTTCATGAAGGTTCAAGTGCTCTAGTTCCTTCCCTCCCGCAGCAAAC 959

QY      115   GACACGGGAGCTGGCCCAAGACCCTCTCTGGGTGA-----TGAGCCCTAAGGGATGGTTTT 167
         || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      960   GACACGGGAGCAGGTCAAGGTGCTCTGGGTAAAAGCCTGTGAGCCTGGGAGTAATCCT 1019

QY      168   GTGPAGGGCCCCRATGCTTCACACTGGGGATCAGACCTCTACCTTCACCCC 218
         || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1020  GTACATGGCTCTTTTACCTACACACTGGGGATTTGACCTCTATCTCCACTC 1070


RESULT 3
US-08-695-191-14
: Sequence 14, Application US/08695191
: Patent No. 6025155
: GENERAL INFORMATION:
: APPLICANT: Hadlaczky, Gyula
: APPLICANT: Szalay, Aladar
: TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
: METHOD OF INVENTION: METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92101-2926
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/695,191
: FILING DATE: 07-AUG-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE: 15-JUL-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/629,822
: FILING DATE: 10-APR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Seidman, Stephanie L.
: REGISTRATION NUMBER: 33,779
: REFERENCE/DOCKET NUMBER: 6869-402C
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-238-0999
: TELEFAX: 619-238-0062
: TELEX:
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1400 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
US-08-695-191-14

Query Match          35.2%; Score 81.4; DB 3; Length 1400;
```

QY 55 GGAAGGCACGCTCTGATTGTCATGAAGGTTCAAGTCTCTAGTTCCTTCCCCCAGGAAAC 114

Sequence 57, Application 05/052300058  
; Patent No. 6458943

; Patent

```

; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210002
; CURRENT APPLICATION NUMBER: US/09/250,609A
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 95
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-609-95

Query Match      12.6%; Score 29.2; DB 4; Length 540;
Best Local Similarity 54.7%; Pred. No. 0.54;
Matches 58; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 26 CAGAGGATAGCTGCTGTTGGCATCTCTGTGGAAGGACGCTGTCATGTCATGAAGGTTTCAG 85
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 CAGAAGCTGTCTAGGAGCTGGCATGATGTGCAGGTCTCTACTGCTTACAGAGACTCAA 286
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 86 TGTCCTAGTTCCCTTCCCTCCAGGAGAAACACACAGCGGAGCTGCCCA 131
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 GAAACTCTTTCACAGGCTGGACAGAAACATCAGCTGCCCTGTCCA 332
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-09-250-611-95
; Sequence 95, Application US/09250611
; Patent No. 6528283
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; APPLICANT: Bassett, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 95
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-611-95

Query Match      12.6%; Score 29.2; DB 4; Length 540;
Best Local Similarity 54.7%; Pred. No. 0.54;
Matches 58; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 26 CAGAGGATAGCTGCTGTTGGCATCTCTGTGGAAGGACGCTGTCATGTCATGAAGGTTTCAG 85
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 CAGAAGCTGTCTAGGAGCTGGCATGATGTGCAGGTCTCTACTGCTTACAGAGACTCAA 286
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 86 TGTCCTAGTTCCCTTCCCTCCAGGAGAAACACACAGCGGAGCTGCCCA 131
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 GAAACTCTTTCACAGGCTGGACAGAAACATCAGCTGCCCTGTCCA 332
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-07-932-454A-2/c
; Sequence 2, Application US/07932454A
; Patent No. 5262318
; GENERAL INFORMATION:
; APPLICANT: GUTHRIE, ELLEN P.
; TITLE OF INVENTION: ISOLATED DNA ENCODING THE Sphi
; TITLE OF INVENTION: RESTRICTION ENDONUCLEASE AND RELATED METHODS FOR PRODUCING
; TITLE OF INVENTION: THE SAME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET

```

```

; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/932,454A
; FILING DATE: 19920820
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: 42078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2692 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 703..1653
; OTHER INFORMATION: /note= "METHYLASE GENE STARTS AT
; OTHER INFORMATION: POSITION 703/ENDS AT 1653. RESTRICTION
; OTHER INFORMATION: ENDONUCLEASE STARTS AT POSITION 1703/ENDS AT 2410"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1703..2410
; US-07-932-454A-2

Query Match      12.4%; Score 28.6; DB 1; Length 2692;
Best Local Similarity 55.6%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 53 GTGGAAGGCAGCTGCTGATTGTCATGAAGGTTTCAGTGTCCCTAGTTCCTTCCCCCAGGAAAA 112
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 418 GGGGGAGTTTCCACCGATTCCATGTCAGTATGTGATGCCATGTGCCCTTCTCTCCAGTACGA 359
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 ACGACACGGGAGCTGCCCAAGACCTCTCTGGGTGATGAG 151
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 GCGGTACGAGCGACGACGACCCCGGCGGTGG 320
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-620-312D-682/c
; Sequence 682, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.

```



```
Db 2196 CACTGAGGCTGCTTCTCCCTGCGGTGAGGACCTGAGTGTGGAGATGATTCCTGCT 2137
QY 143 GGTGATGAGCTAAGGATGTTTGTGTAGGCCCCCTAT---GCTTGACACATGGGGAT 199
Db 2136 TGAATAGTCAGATGAGGCTCCCTTCTGTGTCCTTAGACAGGAGCTCTGGGCT 2077
QY 200 CAGACCTCTACCTTCACCCATGAGGCTT 227
Db 2076 CTAACCTTGGCATCTCTCTGGGAGGTT 2049

RESULT 12
US-08-588-976-21/c
; Sequence 21, Application US/08588976
; Patent No. 5891717
; GENERAL INFORMATION:
; APPLICANT: Christopher B. Newgard, et al.
; TITLE OF INVENTION: Methods and Compositions for
; TITLE OF INVENTION: Inhibiting Hexokinase
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,976
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: UTSD:481/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-588-976-21

Query Match 11.8%; Score 27.2; DB 2; Length 2297;
Best Local Similarity 49.0%; Pred. No. 5.4; Mismatches 103; Indels 3; Gaps 1;
Matches 102; Conservative 0;

QY 23 GCACAGAGGATGAGCTTGTGTCATCCTCTGGAAGGCACGCTGATTCGATGAAGTT 82
Db 2256 GTACAGAGGCTGGCATTTGGGAACGAGGAGAGAGGACCTGGTGTCTTGGAGGA 2197
QY 83 CAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAGGAGCTGGCCACAGACCTCTCTG 142
Db 2196 CACTGAGGGTGGCTTCTTCTCTCTGGTCCAGGACCTGAGTGTGTGAGATGATTCCTGCT 2137
QY 143 GGTGATGAGCTAAGGATGTTTGTGTAGGCCCCCTAT---GCTTGACACATGGGGAT 199
Db 2136 TGAATAGTCAGATGAGGCTCCCTTCTGTGTCCTTAGACAGGAGCTCTGGGCT 2077
QY 200 CAGACCTCTACCTTCACCCATGAGGCTT 227
Db 2076 CTAACCTTGGCATCTCTCTGGGAGGTT 2049
```

```
RESULT 13
5189147-4
; Patent No. 5189147
; APPLICANT: SAITO, HARUO; KRANZ, DAVID M.; EISEN, HERMAN N.;
; TONEGAWA, SUSUMU
; TITLE OF INVENTION: METEORODIMERIC T LYMPHOCYTE RECEPTOR
; ANTIBODY
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/271,216
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 666,988
; FILING DATE: 31-OCT-1984
; APPLICATION NUMBER: 620,122
; FILING DATE: 13-JUN-1984
; SEQ ID NO: 4
; LENGTH: 1054
5189147-4

Query Match 11.7%; Score 27; DB 6; Length 1054;
Best Local Similarity 53.3%; Pred. No. 4.5;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 47 CATCCTGTGGAAGGCACGCTGTGATTCATGAAGGTTGAGTTCAGTTCCTAGTTCCTCCCTCCCA 106
Db 803 CCTACTGGGAAGGCCACCTATATGCTGTGCTGCTGAGTGGCTAGTGTGATGGCCAT 862
QY 107 GGAAGAACGACACGGGAGCTGGCCAGCAAGACCTCTCTGGGTGATGAGCC 153
Db 863 GGTCAAGAAAAAATTCCTGAGACAAAATTTATGTCATCCTGAGCC 909

RESULT 14
US-09-250-609-85
; Sequence 85, Application US/09250609A
; Patent No. 6458943
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210002
; CURRENT APPLICATION NUMBER: US/09/250,609A
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-609-85

Query Match 11.6%; Score 26.8; DB 4; Length 503;
Best Local Similarity 53.9%; Pred. No. 3.9;
Matches 55; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 26 CAGAGGATGATGCTGCTGTGGCATCCTCTGGAAGGCACGCTGATTCATGAAGTTTCA 85
Db 376 CAGAACCTGTCTAGGACCTGGCATGATGTGCAAGTCTCTACTGCTCTACAGNAGACTCA 435
QY 86 TGTCTAGTTCCTTCCCTCCAGGAAAAACGACAGGGAGCTG 127
Db 436 GAAACTCTTTCACAGGCTGGACAGAAAACATCAGCTGCCCTG 477

RESULT 15
US-09-250-611-85
; Sequence 85, Application US/09250611
; Patent No. 6528283
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; APPLICANT: Basset, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
```



```

; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-611-85

```

	Query Match	11.6%	Score 26.8;	DB 4;	Length 503;
	Best Local Similarity	53.9%	Pred. No. 3.9;		
	Matches	55;	Conservative	0;	Mismatches 47; Indels 0; Gaps 0;
QY	26	CAGAGGATACGTGCTCTTGGCATCCTGTGGAGGACGCTCTGATTCGATGAGGTTCA	85		
Db	376	CAGAAACCTGTCTAGGAGCTGGCATGATGTCGAGGTCTCTACTGCTACAAGAGACTCAA	435		
QY	86	TGTCTTCAGTTCCTTCCCCAGGAAAACACACAGGGAGCTG	127		
Db	436	GAAACCTTTTTCAGGCTGGACAGAAACATCATGCTGCCCTG	477		

Search completed: October 9, 2003, 22:44:00  
Job time : 81 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:19:09 ; Search time 3003 Seconds  
(without alignments)  
2551.060 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtaggggtcgaggctaagca.....ttcaccatgaggttgctt 231

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 33363688 seqs, 16581889874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_NA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pna/PCUTUS\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pna/PCUTUS\_COMB.seq.oid:\*
- 3: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq:\*
- 20: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq:\*
- 21: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*
- 22: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*
- 23: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*
- 24: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*
- 25: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*
- 26: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*
- 27: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*
- 28: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*
- 29: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*
- 30: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*
- 31: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*
- 32: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*
- 33: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*
- 34: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq:\*
- 35: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq:\*
- 36: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq:\*
- 37: /cgn2\_6/ptodata/1/pna/US098D\_COMB.seq:\*
- 38: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq:\*
- 39: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq:\*
- 40: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq:\*
- 41: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq:\*
- 42: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq:\*
- 43: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq:\*

44: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq:\*

45: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq:\*

46: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq:\*

47: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq:\*

48: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq:\*

49: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq:\*

50: /cgn2\_6/ptodata/1/pna/US103A\_COMB.seq:\*

51: /cgn2\_6/ptodata/1/pna/US103B\_COMB.seq:\*

52: /cgn2\_6/ptodata/1/pna/US104A\_COMB.seq:\*

53: /cgn2\_6/ptodata/1/pna/US104B\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq:\*

55: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq:\*

56: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq:\*

57: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq:\*

58: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq:\*

59: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq:\*

60: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq:\*

61: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq:\*

62: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq:\*

63: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq:\*

64: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq:\*

65: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq:\*

66: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*

67: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*

68: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq:\*

69: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq:\*

70: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq:\*

71: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq:\*

72: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq:\*

73: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq:\*

74: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq:\*

75: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq:\*

76: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq:\*

77: /cgn2\_6/ptodata/1/pna/US6023A\_COMB.seq:\*

78: /cgn2\_6/ptodata/1/pna/US6023B\_COMB.seq:\*

79: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq:\*

80: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq:\*

81: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq:\*

82: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq:\*

83: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq:\*

84: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq:\*

85: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq:\*

86: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq:\*

87: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq:\*

88: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq:\*

89: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq:\*

90: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq:\*

91: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq:\*

92: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq:\*

93: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq:\*

94: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq:\*

95: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq:\*

96: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq:\*

97: /cgn2\_6/ptodata/1/pna/US6042\_COMB.seq:\*

98: /cgn2\_6/ptodata/1/pna/US6043\_COMB.seq:\*

99: /cgn2\_6/ptodata/1/pna/US6044\_COMB.seq:\*

100: /cgn2\_6/ptodata/1/pna/US6045\_COMB.seq:\*

101: /cgn2\_6/ptodata/1/pna/US6046\_COMB.seq:\*

102: /cgn2\_6/ptodata/1/pna/US6047\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	231	100.0	231	29	US-09-673-716-1	Sequence 1, Appl1
2	231	100.0	391	29	US-09-673-716-4	Sequence 4, Appl1
3	231	100.0	2052	46	US-10-144-771-25944	Sequence 25944, A
4	231	100.0	2052	91	US-60-360-207-25944	Sequence 25944, A



```
RESULT 3
US-10-144-771-25944
; Sequence 25944, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25944
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-25944

Query Match      100.0%; Score 231; DB 46; Length 2052;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1088 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1147

QY 61 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1148 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1207

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
Db 1208 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1267

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 231
Db 1268 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 1318

RESULT 4
US-60-360-207-25944
; Sequence 25944, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25944
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-25944

Query Match      100.0%; Score 231; DB 91; Length 2052;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1088 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1147

QY 61 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1148 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1207

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
Db 1208 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1267

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 231
Db 1268 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 1318
```

```
Db 1268 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 1318

RESULT 5
US-10-144-771-25350
; Sequence 25350, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25350
; LENGTH: 2098
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-25350

Query Match      100.0%; Score 231; DB 46; Length 2098;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1201 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1260

QY 61 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1261 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1320

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
Db 1321 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1380

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 231
Db 1381 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 1431

RESULT 6
US-60-360-207-25350
; Sequence 25350, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25350
; LENGTH: 2098
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-25350

Query Match      100.0%; Score 231; DB 91; Length 2098;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1201 GTGGGTGGAGGCTAAGCACTGCACAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1260

QY 61 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1261 CACGCTCTGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1320

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
Db 1321 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1380

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCCATGAGGCTTGCTT 231
```

```
|||||
Db 1381 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1431

RESULT 7
US-10-144-771-33482
; Sequence 33482, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 33482
; LENGTH: 2258
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-33482

Query Match 100.0%; Score 231; DB 46; Length 2258;
Best Local Similarity 100.0%; Pred. No. 6.2e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 60
Db 1299 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 1358
QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 120
Db 1359 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 1418
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 180
Db 1419 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 1478
QY 181 ATGCTTGCACACTGGGGATCAGACCTTACCTTCACCCATGAGGCTTGCTT 231
Db 1479 ATGCTTGCACACTGGGGATCAGACCTTACCTTCACCCATGAGGCTTGCTT 1529

RESULT 8
US-60-360-207-33482
; Sequence 33482, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 33482
; LENGTH: 2258
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-33482

Query Match 100.0%; Score 231; DB 91; Length 2258;
Best Local Similarity 100.0%; Pred. No. 6.2e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 60
Db 1299 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 1358
QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 120
Db 1359 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 1418
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 180
Db 1419 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 1478
```

```
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1479 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1529

RESULT 9
US-10-144-771-22182
; Sequence 22182, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 22182
; LENGTH: 2353
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-22182

Query Match 100.0%; Score 231; DB 46; Length 2353;
Best Local Similarity 100.0%; Pred. No. 6.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 60
Db 1328 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 1387
QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 120
Db 1388 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 1447
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 180
Db 1448 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 1507
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1508 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1558

RESULT 10
US-60-360-207-22182
; Sequence 22182, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 22182
; LENGTH: 2353
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-22182

Query Match 100.0%; Score 231; DB 91; Length 2353;
Best Local Similarity 100.0%; Pred. No. 6.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 60
Db 1328 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 1387
QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 120
Db 1388 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACAG 1447
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 180
Db 1448 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTTGTGTAGGGCCCT 1507
```

QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231  
|||||  
Db 1508 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1558  
|||||  
RESULT 11  
US-10-144-771-25153  
; Sequence 25153, Application US/10144771  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/10/144,771  
; CURRENT FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 25153  
; LENGTH: 3403  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-10-144-771-25153  
Query Match 100.0%; Score 231; DB 46; Length 3403;  
Best Local Similarity 100.0%; Pred. No. 7e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 60  
|||||  
Db 1537 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 1596  
|||||  
QY 61 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 120  
|||||  
Db 1597 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 1656  
|||||  
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180  
|||||  
Db 1657 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1716  
|||||  
QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231  
|||||  
Db 1717 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1767  
|||||

RESULT 12  
US-60-360-207-25153  
; Sequence 25153, Application US/60360207  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/60/360,207  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 25153  
; LENGTH: 3403  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-360-207-25153  
Query Match 100.0%; Score 231; DB 91; Length 3403;  
Best Local Similarity 100.0%; Pred. No. 7e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 60  
|||||  
Db 1537 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 1596  
|||||  
QY 61 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 120  
|||||  
Db 1597 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 1656  
|||||  
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180  
|||||

Db 1657 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1716  
QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231  
|||||  
Db 1717 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1767  
|||||  
RESULT 13  
US-10-144-771-28382  
; Sequence 28382, Application US/10144771  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/10/144,771  
; CURRENT FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 28382  
; LENGTH: 5633  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-10-144-771-28382  
Query Match 100.0%; Score 231; DB 46; Length 5633;  
Best Local Similarity 100.0%; Pred. No. 8.3e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 60  
|||||  
Db 1641 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 1700  
|||||  
QY 61 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 120  
|||||  
Db 1701 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 1760  
|||||  
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180  
|||||  
Db 1761 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1820  
|||||  
QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231  
|||||  
Db 1821 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1871  
|||||

RESULT 14  
US-60-360-207-28382  
; Sequence 28382, Application US/60360207  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/60/360,207  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 28382  
; LENGTH: 5633  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-360-207-28382  
Query Match 100.0%; Score 231; DB 91; Length 5633;  
Best Local Similarity 100.0%; Pred. No. 8.3e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 60  
|||||  
Db 1641 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGAAGG 1700  
|||||  
QY 61 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 120  
|||||  
Db 1701 CAGCTCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAG 1760  
|||||  
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180  
|||||





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:41:03 ; Search time 171 Seconds  
(without alignments)  
1886.736 Million cell updates/sec

Title: US-09-673-716-1  
 Perfect score: 231  
 Sequence: 1 gtaggggtgcaggctaagca.....ttcacccatgaggttgctt 231

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1776535 seqs, 698337343 residues

Total number of hits satisfying chosen parameters: 3553070

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	72	31.2	96593	1	PCT-US02-38582-151	Sequence 151, App
2	35	15.2	30593	7	US-60-495-114-16231	Sequence 16231, A
C 3	34	14.7	49596	7	US-60-495-135-3644	Sequence 3644, Ap
4	32.6	14.1	1106	6	US-10-425-114A-16370	Sequence 16270, A
5	32.6	14.1	1145	6	US-10-425-114A-25489	Sequence 25489, A
6	31	13.4	1205	6	US-10-425-114A-16110	Sequence 16110, A
7	31	13.4	1211	6	US-10-425-114A-23937	Sequence 23937, A
8	30.2	13.1	25995	7	US-60-500-337-19598	Sequence 19598, A
9	29.8	12.9	37637	7	US-60-500-315-11368	Sequence 11368, A
10	29.6	12.8	959	6	US-10-663-561-452	Sequence 452, App
11	29.6	12.8	1531	6	US-10-425-114A-1045	Sequence 1045, Ap
C 12	29.6	12.8	2731748	6	US-10-297-4659B-1	Sequence 1, Appli
13	29.4	12.7	19946	7	US-60-500-337-19711	Sequence 19711, A
14	28.8	12.5	34227	7	US-60-500-315-11594	Sequence 11594, A
C 15	28.4	12.3	25821	7	US-60-487-610-20089	Sequence 20089, A
C 16	28.4	12.3	25821	7	US-60-485-450-12348	Sequence 12348, A
C 17	28.2	12.2	15448	7	US-60-500-315-11760	Sequence 11760, A
C 18	27.8	12.0	201	7	US-60-485-450-10495	Sequence 10495, A
19	27.8	12.0	201	7	US-60-485-450-10505	Sequence 10505, A
20	27.8	12.0	201	7	US-60-485-450-10515	Sequence 10515, A
21	27.8	12.0	201	7	US-60-485-450-41518	Sequence 41518, A
22	27.8	12.0	34570	1	PCT-US02-38582-157	Sequence 157, App
23	27.8	12.0	45945	7	US-60-500-337-19316	Sequence 19316, A
C 24	27.8	12.0	85568	7	US-60-495-114-16789	Sequence 16789, A
C 25	27.8	12.0	85568	7	US-60-495-135-3707	Sequence 3707, Ap
26	27.8	12.0	329776	5	US-09-947-914-47	Sequence 49, Appli

## ALIGNMENTS

```

RESULT 1
PCT-US02-38582-151
; Sequence 151, Application PC/TUS0238582
; GENERAL INFORMATION:
; APPLICANT: SAGRES DISCOVERY
; APPLICANT: MORRIS, David W.
; APPLICANT: ENGELHARD, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000143
; CURRENT APPLICATION NUMBER: PCT/TUS02/38582
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/997,722
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151
; LENGTH: 96593
; TYPE: DNA
; ORGANISM: Mus Musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(96593)
; OTHER INFORMATION: n = A,T,C or G
; PC-US02-38582-151

```

	Query Match	31.2%	Score 72	DB 1	Length 96593
	Best Local Similarity	72.1%	Pred. No. 6.7e-14		
	Matches 124	Conservative 0	Mismatches 40	Indels 8	Gaps 2
QY	55	GGAAGGCAGCTCTGATTGATGAGCTTCAGTGTCTTACGTCCTCCCTTC-CCCACGGAGAAA	113		
Db	1044	GGGAGACATGTCATCTTTCAAGAAGGTGTAGTGTCCAAGTGTCTTCTCCACGGCAAAA	1103		
QY	114	CGACACGGGAGCTTGGCCAAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGGTTT	166		
Db	1104	CGACACGGGACAGGTTCAGGGTGTCTCTGGGTAAAGCCTGTGAGCCTAAGAGCTAATCC	1163		
QY	167	TGRTGAGGGCCCCCTATGCTTTCACATCTGGGGATCAGACCTCTACCTTCACCC	218		
Db	1164	TGTACATGGCTCTTTTACCTGTGCACATCGGGGATTTGACCTCTATCTCCACTC	1215		

RESULT 2  
US-60-495-114-16231  
; Sequence 16231, Application US/60495114  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL001480  
; CURRENT APPLICATION NUMBER: US/60/495,114  
; CURRENT FILING DATE: 2003-08-15  
; NUMBER OF SEQ ID NOS: 91238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16231  
; LENGTH: 30593  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-60-495-114-16231

Query Match 15.2%; Score 35; DB 7; Length 30593;  
Best Local Similarity 55.3%; Pred. No. 0.21;  
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 4 GGGTGGCAGGTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGGAAGGCAC 63  
DB 15181 GGGTGGGGACACTCAGAGCCACAGAGACTCAGAGCTCTGTGCCATGGCCCAAGCCCT 15240  
QY 64 GTCTGATTGCATGAAGTTTCAGTCTCCTAGTTCCTTCCCGAGGAAAAAGACACGGGA 123  
DB 15241 GTTGGATGTTGGCATCTGCCCTGCCAGGACCCCGCCAGCCACAGGACAGAG 15300  
QY 124 GCT 126  
DB 15301 GCT 15303

RESULT 3  
US-60-495-135-3644/c  
; Sequence 3644, Application US/60495135  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL001479  
; CURRENT APPLICATION NUMBER: US/60/495,135  
; CURRENT FILING DATE: 2003-08-15  
; NUMBER OF SEQ ID NOS: 18339  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3644  
; LENGTH: 49596  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-60-495-135-3644

Query Match 14.7%; Score 34; DB 7; Length 49596;  
Best Local Similarity 57.5%; Pred. No. 0.52;  
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 11 AGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGGAAGGCACGTCATGAT 70  
DB 4553 AGGCAGACTCGGACACTATCTGTGTACCGCATATTCAGCTGGGACTCTTAC 4494  
QY 71 TGCATGAAGTTTCAGTCTCCTAGTTCCTTCCCGAGGAAAAAGGA 116  
DB 4493 TTCAGCCCTTCTCCTGTGCCAGTTCCTTGGCAGGACATCCA 4448

RESULT 4  
US-10-425-114A-16270  
; Sequence 16270, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 16270  
; LENGTH: 1106  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3062-025-A9\_FLI  
US-10-425-114A-16270

Query Match 14.1%; Score 32.6; DB 6; Length 1106;  
Best Local Similarity 52.6%; Pred. No. 0.51;  
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 31 GATAGCTTGTGTGGCATCTCTGGAAGGCACGCTCTGATTCATGAAGGTTCAAGTGTC 90  
DB 715 GCTAGAATGGTCCAGCCCAACGACGCGGAAGCTGGTATGGTTTCTGGACAAGCGCGGCA 774  
QY 91 TAGTTCCTTCCCGAGGAAAAAGACACGCGGAGCTGCCCAAGACCTCTCTGGGTGATGA 150  
DB 775 TCATTCTTGTACTCCAGACGACGACGACGACGCTTGTACGATCGTCACAGTGACTA 834  
QY 151 GCCTAAGGGATGGTT 165  
DB 835 ACTCTGAGGATCGTT 849

RESULT 5  
US-10-425-114A-25489  
; Sequence 25489, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 25489  
; LENGTH: 1145  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3912-015-C11\_FLI  
US-10-425-114A-25489

Query Match 14.1%; Score 32.6; DB 6; Length 1145;  
Best Local Similarity 52.6%; Pred. No. 0.52;  
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 31 GATAGCTTGTGTGGCATCTCTGGAAGGCACGCTCTGATTCATGAAGGTTCAAGTGTC 90  
DB 753 GCTAGAATGGTCCAGCCCAACGACGCGGAAGCTGGTATGGTTTCTGGACAAGCGCGGCA 812  
QY 91 TAGTTCCTTCCCGAGGAAAAAGACACGCGGAGCTGCCCAAGACCTCTCTGGGTGATGA 150  
DB 813 TCATTCTTGTACTCCAGACGACGACGACGACGCTTGTACGATCGTCACAGTGACTA 872  
QY 151 GCCTAAGGGATGGTT 165  
DB 873 ACTCTGAGGATCGTT 887

RESULT 6  
US-10-425-114A-16110

; Sequence 16110, Application US/10425114A  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114A

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 16110

; LENGTH: 1205

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3061-082-A12\_FLI

US-10-425-114A-16110

Query Match 13.4%; Score 31; DB 6; Length 1205;

Best Local Similarity 51.9%; Pred. No. 1.8;

Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 31 GATAGCTTGCTTGGCATCTCTGGTGAAGGCGACGCTGATTCATGCAAGGTTTCAGTGTCC 90

DB 815 GCTAGAAATGGTCCAGCCACGCGGAAGCTGGTATGTTTCTGGACAAGCGCGGCA 874

QY 91 TAGTTCCTTCCCCCAGGAAAACGACACGGGAGCTGGCCAGACCTCTCTGGGTGTGA 150

DB 875 TCATTCTTGTGCTCAAGAACGACGATGACGACGCTTGTACGAGCGTCATGATGACTA 934

QY 151 GCTAAGGAGGTGTT 165

DB 935 ACTCGAGGATCCTT 949

#### RESULT 7

US-10-425-114A-23937

; Sequence 23937, Application US/10425114A

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114A

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 23937

; LENGTH: 1211

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3606-028-C5\_FLI

US-10-425-114A-23937

Query Match 13.4%; Score 31; DB 6; Length 1211;

Best Local Similarity 51.9%; Pred. No. 1.8;

Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 31 GATAGCTTGCTTGGCATCTCTGGTGAAGGCGACGCTGATTCATGCAAGGTTTCAGTGTCC 90

DB 787 GCTAGAAATGGTCCAGCCACGCGGAAGCTGGTATGTTTCTGGACAAGCGCGGCA 846

QY 91 TAGTTCCTTCCCCCAGGAAAACGACACGGGAGCTGGCCAGACCTCTCTGGGTGTGA 150

DB 875 TCATTCTTGTGCTCAAGAACGACGATGACGACGCTTGTACGAGCGTCATGATGACTA 934

DB 847 TCATTCCTTGACTCCAAAGAACGACGATGACGAGCTTGTACGAGCGTCATGATGACTA 906

QY 151 GCCTAAGGAGGTGTT 165

DB 907 ACTCCGAGGATCCTT 921

#### RESULT 8

US-60-500-337-19598

; Sequence 19598, Application US/60500337

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND

; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: CL001483

; CURRENT APPLICATION NUMBER: US/60/500,337

; CURRENT FILING DATE: 2003-09-05

; NUMBER OF SEQ ID NOS: 123188

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19598

; LENGTH: 25995

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-500-337-19598

Query Match 13.1%; Score 30.2; DB 7; Length 25995;

Best Local Similarity 53.9%; Pred. No. 8.6;

Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 59 GGCAGCTGTGATGATGAAGTTCAGTGTCTCAGTTCCTTCCCTCCCGCAAGAAACGACA 118

DB 3083 GGCAGGTGCGAGGGGCTCAGGGTTCACCTCTGCTGGCTCGGCTGACCCCTCTGGCTCTCA 3142

QY 119 CGSGAGCTGCGCAAGACCTCTCTGGTGATGACCTTAAGGATGTTTTGTGTAG 173

DB 3143 CAGGCGCTGGACTGCGCGGCTGGTGGGGAGGCGGAGGACTCGGAGGGGGAAG 3197

#### RESULT 9

US-60-500-315-11368

; Sequence 11368, Application US/60500315

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS

; TITLE OF INVENTION: OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001484

; CURRENT APPLICATION NUMBER: US/60/500,315

; CURRENT FILING DATE: 2003-09-05

; NUMBER OF SEQ ID NOS: 69978

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11368

; LENGTH: 37637

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-500-315-11368

Query Match 12.9%; Score 29.8; DB 7; Length 37637;

Best Local Similarity 63.0%; Pred. No. 13;

Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 51 CTGTGGAAGGCGACGCTGATTCATGCAAGGTTTCAGTGTCTCCTAGTTCCTTCCCTCCCGCAGGAA 110

DB 36010 CTCCTGGGAGGAGCTCTGTTTCAATGTGAGGCTTGTGTATATTTCCCTTCACACAGGA 36069

QY 111 AAACGACACGGGA 123

DB 36070 TCACGTCACGTGA 36082

#### RESULT 10

US-10-663-561-452

```
; Sequence 452, Application US/10663561
; GENERAL INFORMATION:
; APPLICANT: DENSLON, NANCY D.
; APPLICANT: LARKIN, PATRICK M.
; APPLICANT: FOLMAR, LEROY C.
; APPLICANT: HEMMER, MICHAEL J.
; TITLE OF INVENTION: DETECTING HORMONALLY ACTIVE COMPOUNDS
; FILE REFERENCE: 5853-238
; CURRENT APPLICATION NUMBER: US/10/663,561
; CURRENT FILING DATE: 2003-09-15
; NUMBER OF SEQ ID NOS: 566
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 452
; TYPE: DNA
; ORGANISM: Micropterus salmoides
US-10-663-561-452

Query Match      12.8%; Score 29.6; DB 6; Length 959;
Best Local Similarity 56.0%; Pred. No. 5.2;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 103 CCCAGAAAAACGACACGGGAGCTCTCTGGGTGATGAGCTAAGGGATG 162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 CTCAGAAAAAAGAGAGATGTGGAGCTGTGGACCGAGAGCGCTGAGCCAAAGTTGT 479

QY 163 GTTTTGTAGGCCCTATGCTTGCACACTGGGGATCAG 202
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 GTTTCTGTGGACCTCTGATCATGATGCGGCTTCTGAACAG 519

RESULT 11
US-10-425-114A-1045
; Sequence 1045, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1045
; LENGTH: 1531
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700100671_FLI
US-10-425-114A-1045

Query Match      12.8%; Score 29.6; DB 6; Length 1531;
Best Local Similarity 54.6%; Pred. No. 5.9;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 56 GAAGGCACGCTCATTCGATGAGGTTTCAGTGTCTCTCCCTTCCCCCAGGAAAACG 115
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 157 GAAGTCCATGCTACCCCAATTAAGTCCGAGCGTCCGCAACAGGCTCCGCGCAGAGACG 216

QY 116 ACACGGGAGTGCCCAAGACCTCTCTGGGTGATGAGCTAAGGGATGG 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 AATGCGGAGNATGGACGACGAGAGGTTGATGACCCAAAGGAATTG 264

RESULT 12
US-10-297-465B-1
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
```

```
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

Query Match      12.8%; Score 29.6; DB 6; Length 2731748;
Best Local Similarity 64.7%; Pred. No. 52;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 68 GATTGATGAAGTTTCAGTGTCTCTAGTTCCTTCCCCAGAAAAACGACACGGGAGCTG 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 719549 GTTTGGATGAACCTCCAGAGTGTCTGTCGCTTCCGCGGAAAAATCGGCGCGGCTT 719608

QY 128 GCCAAGAC 135
      |||||
Db 719609 CCCGCGTC 719616

RESULT 13
US-60-500-337-19711/c
; Sequence 19711, Application US/60500337
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: US$ THEREOF
; FILE REFERENCE: CL001483
; CURRENT APPLICATION NUMBER: US/60/500,337
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 123188
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19711
; LENGTH: 19946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-337-19711

Query Match      12.7%; Score 29.4; DB 7; Length 19946;
Best Local Similarity 54.1%; Pred. No. 15;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 67 TGATTGCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCCAGAAAAACGACACGGGAGCT 126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 662 TCAGTCGGGAAAGGGCGGCTCGAGGCTCCCTCCCCAGCCCCCACATCTGTGGGCT 603

QY 127 GGCCAGAGACTCTCTCTGGGTGATGAGCTAAGGATGTTTGTGTAGGGCC 177
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 602 GGCCCCAGCATAGCTGGGAGGAGCAGCTGTGGTCTTGTCTGAGCTCGTGAC 552

RESULT 14
US-60-500-315-11594
; Sequence 11594, Application US/60500315
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS
; TITLE OF INVENTION: OF DETECTION AND USES THEREOF
```

```
; FILE REFERENCE: CL001484
; CURRENT APPLICATION NUMBER: US/60/500,315
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 69978
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11594
; LENGTH: 34227
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-315-11594

Query Match      12.5%; Score 28.8; DB 7; Length 34227;
Best Local Similarity 58.0%; Pred. No. 28;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 115 GACAGGGGAGGTGGCCAGAGCTCTCTCGGGTGATGAGCCTAAGGGGATGGTTTTGTGTAGG 174
Db 17633 GAGGCTCGGGCTGCGCAGGAGCCACAGAGGGGGAGGCTAAGGAATGGCAGGCTGCAGG 17692

QY 175 GCCCTATGCTTGCACACTGGGGATCAG 202
Db 17693 TCCCAAGCCCTGCCCCACAGGGAGGCAG 17720

RESULT 15
US-60-487-610-20089/c
; Sequence 20089, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20089
; LENGTH: 25821
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-20089

Query Match      12.3%; Score 28.4; DB 7; Length 25821;
Best Local Similarity 58.1%; Pred. No. 35;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 51 CTGTGGAGGCACGCTCTGATTGCATGAAGGTTTCAGTCTCCTAGTTCCTTCCCCCAGGAA 110
Db 22839 CTAGGGAATGCTGGTTCACATCGGTGGAAGGTTAGGGTCTGGGGGAAGCTGGCAGCAGCA 22780

QY 111 AAACACACACGGGAGCTGGCCAAAGCC 136
Db 22779 GATGCACACCGGACCTGGTCAAGGCC 22754

Search completed: October 9, 2003, 23:37:22
Job time : 179 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 . Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 20:13:18 ; Search time 256 Seconds  
(without alignments)  
2435.821 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 ggggggctggaggaagca.....ttcaccatgaggctgtgctt 231

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231	100.0	231	21	AAZ55983
2	231	100.0	393	21	AAZ55986
3	224	97.0	226	21	AAZ55987
4	114.4	49.5	75798	25	ABX77212
5	89.4	38.7	1870	25	ABZ80240
6	89.4	38.7	5018	6	AAZ50150
7	89.4	38.7	5018	8	AAZ70646
8	87.8	38.0	3336	6	AAZ50149
					Core fragment B, a murine genomic pos
					Murine genomic pos
					Post-transcription
					Mouse uridine diph
					Mouse tramdorin 3
					Sequence of the cd
					Clone 23Bp102 enc
					Sequence of the cd

9	86.2	37.3	1434	18	AAV04900
10	82	35.5	1434	25	ABX11082
11	81.4	35.2	1400	18	AAV04901
12	78.6	34.0	1400	25	ABX11083
13	35	15.2	50	16	AAQ83497
14	34	14.7	122186	22	AAZ89560
15	30.4	13.2	475	24	ABL81597
16	30.4	13.2	691	21	AAZ01644
17	30	13.0	30	21	AAZ55984
18	30	13.0	610	20	AAZ61423
19	29.4	12.7	32187	22	AAZ32249
20	29.4	12.7	32187	24	ABZ67552
21	29.2	12.6	540	20	AAZ08623
22	29.2	12.6	540	21	AAZ35795
23	29.2	12.6	1111	22	AAZ94634
24	29.2	12.6	2648	25	ACC46566
25	29.2	12.6	3881	22	AAI69907
26	29.2	12.6	5176	25	ACC46562
27	29.2	12.6	9040	24	ABL40559
28	29.2	12.6	9154	24	ABZ71695
29	29.2	12.6	9418	24	ABZ71693
30	29.2	12.6	9894	24	ABZ71691
31	29.2	12.6	9993	24	ABZ71690
32	29.2	12.6	9993	24	ABZ71692
33	29.2	12.6	14834	24	ABZ83570
34	29	12.6	820	21	AAZ51239
35	29	12.6	823	21	AAZ51535
36	28.8	12.5	663	22	AAZ33056
37	28.8	12.5	664	22	AAZ33280
38	28.8	12.5	703	21	AAZ01665
39	28.6	12.4	427	23	ABZ42209
40	28.6	12.4	2529	21	AAZ29467
41	28.6	12.4	2588	21	AAZ29469
42	28.6	12.4	2692	14	AAQ51114
43	28.6	12.4	3721	23	ABL07213
44	28.6	12.4	6465	23	ABL07212
45	28.6	12.4	6549	24	AAZ594790

#### ALIGNMENTS

##### RESULT 1

AAZ55983  
ID AAZ55983 standard; DNA; 231 BP.  
XX  
AC AAZ55983;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Core fragment B, a murine genomic core PRE region.  
XX  
KW Post translational regulatory element; PRE; murine; NCTE;  
KW post-transcriptional RNA nucleo-cytoplasmic transport element;  
KW attenuated virus; retrovirus; HIV-1; intracisternal A-type particle; IAP;  
KW vaccine; viral infection; core fragment B; ds.  
XX  
OS Mus sp.  
XX  
PN WO9961596-A2.  
XX  
XX  
PD 02-DEC-1999.  
XX  
PF 18-MAY-1999; 99WO-US11082.  
XX  
PR 22-MAY-1998; 98US-0086487.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Pavlakis GN, Nappi F;  
XX  
DR WPI; 2000-116336/10.  
XX

Nucleotide sequenc  
DNA sequence of ne  
Nucleotide sequenc  
DNA sequence of ne  
DNA probe 3 detect  
Human histone deac  
Human ovarian canc  
Human colon cancer  
Murine post transl  
DNA encoding a hum  
Human DNA repair a  
Novel human DNA re  
EST with homology  
Expressed sequence  
Human full-length  
Human dithp secret  
Human fibrillin 3  
Human dithp secret  
Human fibrillin 3  
DNA encoding human  
DNA encoding human  
DNA encoding human  
DNA encoding human  
Human cDNA differe  
Arabidopsis thalia  
Arabidopsis thalia  
DNA encoding human  
DNA encoding human  
Human colon cancer  
Genomic sequence #  
P. chrysogenum sut  
Genomic P. chrysog  
Sequence encoding  
Drosophila melanog  
Drosophila melanog  
Human DNA sequence

PT Novel post-transcriptional regulatory elements used to construct  
 XX attenuated retroviruses for vaccines  
 PS Claim 3: Page 44; 59pp; English.  
 XX This sequence represents a core post-transcriptional regulatory  
 CC element (PRE) region, designated core fragment B, which is located  
 CC within PRE7 (AAZ55986). Core fragment B can function as a post-  
 CC transcriptional RNA nucleocytoplasmic transport element (NCTE). The PRE  
 CC is used to replace the NCTE of HIV-1, producing a virus with lower  
 CC replicative activity, resulting in an attenuated virus. This novel PRE  
 CC was initially derived from a murine genomic sequence and sequence  
 CC analysis found that it had significant homology to intracisternal A-type  
 CC particle (IAP) sequences. The nucleic acids and oligonucleotides of the  
 CC invention can be delivered into cells cultures, tissues and organisms  
 CC for synthesis, mutation and screening. When the PRE is used to replace  
 CC the NCTE of viruses, especially retroviruses such as HIV, an attenuated  
 CC virus is produced, which may be used in a viral vaccine for the  
 CC prophylaxis or amelioration of a viral infection in a mammal.  
 XX Sequence 231 BP; 47 A; 58 C; 70 G; 56 T; 0 other;  
 SQ Query Match 100.0%; Score 231; DB 21; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 3e-70;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTTGTGGCATCTGTGGAAGG 60  
 DB 1 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTTGTGGCATCTGTGGAAGG 60  
 QY 61 CAGCTCTGATTGATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCCGAGAAAACGACAG 120  
 DB 61 CAGCTCTGATTGATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCCGAGAAAACGACAG 120  
 QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
 DB 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
 QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231  
 DB 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231  
 RESULT 2  
 AAZ55986  
 ID AAZ55986 standard; DNA; 393 BP.  
 AC AAZ55986;  
 XX  
 XX 10-APR-2000 (first entry)  
 XX Murine genomic post-transcriptional regulatory element, PRE7.  
 DE  
 XX Post translational regulatory element; PRE; murine; NCTE;  
 KW post-transcriptional RNA nucleocytoplasmic transport element;  
 KW attenuated virus; retrovirus; HIV-1; Intracisternal A-type particle; IAP;  
 KW vaccine; viral infection; PRE7; ds.  
 XX Mus sp.  
 XX WO9961596-A2.  
 XX  
 XX 02-DEC-1999.  
 XX 18-MAY-1999; 99WO-US11082.  
 XX 22-MAY-1998; 98US-0086487.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Pavlakis GN, Nappi F;  
 XX WPI; 2000-116336/10.  
 DR

XX Novel post-transcriptional regulatory elements used to construct  
 PT attenuated retroviruses for vaccines  
 PS Example 1; Page 44; 59pp; English.  
 XX This invention relates to the core region of a novel post-transcriptional  
 CC regulatory element (PRE), designated core fragment B (AAZ55983), that  
 CC can function as a post-transcriptional RNA nucleocytoplasmic transport  
 CC element (NCTE). The PRE is used to replace the NCTE of HIV-1, producing  
 CC a virus with lower replicative activity, resulting in an attenuated  
 CC virus. This novel PRE was initially derived from a murine genomic  
 CC sequence and sequence analysis found that it had significant homology to  
 CC intracisternal A-type particle (IAP) sequences. The nucleic acids and  
 CC oligonucleotides of the invention can be delivered into cells cultures,  
 CC tissues and organisms for synthesis, mutation and screening. When the  
 CC PRE is used to replace the NCTE of viruses, especially retroviruses such  
 CC as HIV, an attenuated virus is produced, which may be used in a viral  
 CC vaccine for the prophylaxis or amelioration of a viral infection in a  
 CC mammal. This sequence represents PRE7, a murine sequence which comprises  
 CC core fragment B which was characterised in an exemplification of the  
 CC present invention, and is functional as a PRE.  
 XX Sequence 393 BP; 85 A; 97 C; 109 G; 102 T; 0 other;  
 SQ Query Match 100.0%; Score 231; DB 21; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-70;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTTGTGGCATCTGTGGAAGG 60  
 DB 99 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTTGTGGCATCTGTGGAAGG 158  
 QY 61 CAGCTCTGATTGATGAAGTTTCAGTGTCTCTAGTTCCTTCCCGAGAAAACGACAG 120  
 DB 159 CAGCTCTGATTGATGAAGTTTCAGTGTCTCTAGTTCCTTCCCGAGAAAACGACAG 218  
 QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180  
 DB 219 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 278  
 QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231  
 DB 279 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 329  
 RESULT 3  
 AAZ55987  
 ID AAZ55987 standard; DNA; 226 BP.  
 AC AAZ55987;  
 XX  
 XX 10-APR-2000 (first entry)  
 XX Post-transcriptional regulatory element PRE7 functional fragment, M4.  
 DE  
 XX Post translational regulatory element; PRE; murine; NCTE;  
 KW post-transcriptional RNA nucleocytoplasmic transport element;  
 KW attenuated virus; retrovirus; HIV-1; Intracisternal A-type particle; IAP;  
 KW vaccine; viral infection; M4; ds.  
 XX Unidentified.  
 XX WO9961596-A2.  
 XX  
 XX 02-DEC-1999.  
 XX 18-MAY-1999; 99WO-US11082.  
 XX 22-MAY-1998; 98US-0086487.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX







```

XX FH Key Location/Qualifiers
XX FT CDS 423..2753
XX FT /*tag= a
XX PN JP62045600-A.
XX PD 27-FEB-1987.
XX PF 21-AUG-1985; 85JP-0183810.
XX PR 21-AUG-1985; 85JP-0183810.
XX PA (SCHE ) SCHERING BIOTECH CORP.
XX DR WPI; 1987-097765/14.
XX DR P-PSDB; AAP70417.
XX PT cDNA clone coding polypeptide - showing IgE bond factor activity.
XX PS Disclosure; Page 29; 33pp; Japanese.
XX CC The clone encodes a polypeptide which shows IgE binding factor
XX CC activity. See also AAN70645.
XX CC (Updated on 10-MAR-2003 to add missing OS field.)
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 5018 BP; 1423 A; 1087 C; 1165 G; 1343 T; 0 other;
Query Match 38.7%; Score 89.4; DB 8; Length 5018;
Best Local Similarity 74.9%; Pred. No. 1.6e-20;
Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;
QY 55 GGAAGGCACGCTGATTGCATGAAGGTTTCAGTGTCTAGTTCCTTCCCCAGGAAAAAC 114
DB 4449 GGGAGACATGTCATCTTTTCATGAAGTTTCAGTGTCTAGTTCCTTCCCCAGGAAAAAC 4508
QY 115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCTTAAGGGATGGTTT 167
DB 4509 GACACGGGAGCAGGTCTGGGTGCTCTGGGTAAAAGCCTGTGAAGCCTAAGAGCTAATCCT 4568
QY 168 GTGTAGGCGCCCTATGCTTGCACACTGGGGATCAGACCTCTACCTTTCACCC 218
DB 4569 GTACATGGCTCCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 4619

RESULT 8
AAN50149
ID AAN50149 standard; cDNA; 3336 BP.
XX AC AAN50149;
XX DT 25-MAR-2003 (updated)
XX DT 16-AUG-2002 (updated)
XX DT 27-SEP-1991 (first entry)
XX DE Sequence of the cDNA clone 23b6p8.3 encoding a polypeptide
XX DE exhibiting mammalian immunoglobulin binding factor activity (IBF).
XX KW Immunoglobulin E-mediated disease; therapy; B-cell differentiation;
XX KW immunoglobulin E-binding factor; ss.
XX OS Rattus sp.
XX FH Key Location/Qualifiers
XX FT CDS 94..1767
XX FT /*tag= a
XX PN EP155192-A.
XX PD 18-SEP-1985.
XX PF 15-MAR-1985; 85EP-0301834.

```

```

XX PR 16-MAR-1984; 84US-0590430.
XX PA (SCHE ) SCHERING BIOTECH CORP.
XX PA (UYJO ) UNIV JOHNS HOPKINS.
XX PA (SCHE ) SCHERING-BIOTECH CO.
XX PA (DNAX-) DNAX RES INST MOLEC.
XX PI Martens CL, Ishizaka K, Moore KW, Huff TF;
XX DR WPI; 1985-231863/38.
XX DR P-PSDB; AAP50120.
XX PT New complementary DNA clones coding for poly:peptide(s) - with
XX PT sequence of mammalian immunoglobulin factor and obtd. from
XX PT transformed or transfected host
XX PS Claim 8; Page 55-58; 71pp; English.
XX CC IBF and IBF cDNA are useful in studies on the immune system.
XX CC Treatment for Igb-mediated diseases may be possible. IBF may enhance
XX CC B-cell differentiation into an immunoglobulin-secreting cell.
XX CC (Updated on 16-AUG-2002 to add missing OS field.)
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 3336 BP; 944 A; 725 C; 839 G; 828 T; 0 other;
Query Match 38.0%; Score 87.8; DB 6; Length 3336;
Best Local Similarity 74.3%; Pred. No. 5.1e-20;
Matches 127; Conservative 0; Mismatches 37; Indels 7; Gaps 1;
QY 55 GGAAGGCACGCTGATTGCATGAAGGTTTCAGTGTCTAGTTCCTTCCCCAGGAAAAAC 114
DB 2772 GGGAAACATGTCATCTTTTCATGAAGTTTCAGTGTCTAGTTCCTTCCCCAGGAAAAAC 2831
QY 115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCTTAAGGGATGGTTT 167
DB 2832 GACACGGGAGCAGGTCTGGGTGCTCTGGGTAAAAGCCTGTGAAGCCTAAGAGCTAATCCT 2891
QY 168 GTGTAGGCGCCCTATGCTTGCACACTGGGGATCAGACCTCTACCTTTCACCC 218
DB 2892 GTACATGGCTCCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 2942

RESULT 9
AAV04900
ID AAV04900 standard; DNA; 1434 BP.
XX AC AAV04900;
XX DT 30-APR-1998 (first entry)
XX DE Nucleotide sequence of a portion of a neo-minichromosome.
XX KW Neo-minichromosome; multicentric; dicentric; chromosome; euchromatin;
XX KW heterochromatic DNA; minichromosome; artificial chromosome;
XX KW selectable marker; satellite artificial chromosome; SATAC; genomic locus;
XX KW targeted integration; transgenic animal; therapeutic product;
XX KW gene therapy; cloning vehicle; genomic DNA library; ss.
XX OS Mus sp.
XX PN WO9740183-A2.
XX PD 30-OCT-1997.
XX PF 10-APR-1997; 97WO-US05911.
XX PR 07-AUG-1996; 96US-0695191.
XX PR 10-APR-1996; 96US-0629822.
XX PR 15-JUL-1996; 96US-0682080.
XX PA (AMGE-) AMERICAN GENE THERAPY INC.

```

[illegible]

KW	XX	bacteria; plasmid; neo-minichromosome; EC3/7C5 cell; ds.
OS	XX	Homo sapiens.
OS	XX	Mus sp.
OS	XX	Bacteriophage lambda.
OS	XX	Bacteriaceae.
OS	XX	Synthetic.
FT	XX	Key
FT	XX	Location/Qualifiers
FT	XX	misc_feature 1..1434
FT	XX	/*tag= a
FT	XX	/note= "All N nucleotides have been added by the
FT	XX	indexer to match the number of bases this
FT	XX	sequence is listed to contain in the
FT	XX	sequence listing"
XX	XX	
PN	XX	US2002160410-A1.
PD	XX	31-OCT-2002.
XX	XX	17-APR-2002; 2002US-0125767.
PF	XX	28-NOV-2000; 2000US-0724693.
XX	XX	10-APR-1997; 97US-0835682.
PR	XX	10-APR-1996; 96US-0629822.
PR	XX	15-JUL-1996; 96US-0682080.
PR	XX	07-AUG-1996; 96US-0695191.
XX	XX	(HADL/) HADLACZKY G.
PA	XX	(SZAL/) SZALAY A A.
XX	XX	Hadlaczky G, Szalay AA;
PI	XX	WPI; 2003-265757/26..
XX	XX	Amplifying nucleic acid for constructing artificial chromosomes, comprises introducing a DNA fragment into a cell to incorporate the DNA into a chromosome, and identifying cells with chromosomes that has undergone amplification -
XX	XX	Example 2; Page 63; 107pp; English.
XX	XX	The present invention relates to methods for amplifying nucleic acids in cells, for isolating artificial chromosomes, and preparing cell lines that contain artificial chromosomes. The methods comprise introducing a DNA fragment comprising a selectable marker into a cell, growing the cell under selective conditions to produce cells that have incorporated the DNA fragment or its portion into a chromosome, and identifying from among the resulting cells those that include a chromosome or its fragment with a portion that has undergone amplification. The invention also discloses cells and cell lines that contain the nucleic acids or artificial chromosomes. The methods of the invention are useful for amplifying nucleic acids in cells, in generating and isolating artificial chromosomes (e.g. mammalian artificial chromosomes or MACs), and in delivering the chromosomes to selected cells and tissues. The artificial chromosomes are useful in gene therapy, gene product production systems, production of humanised genetically transformed animal organs, production of transgenic plants and animals that would employ chromosomal elements as information storage vehicles, for analysis and study of centromere function, for the production of artificial chromosome vectors, and for the preparation of species-specific artificial chromosomes. The present sequence represents a neo-minichromosome (composed of human, bacteriophage lambda, plasmid, bacterial and mouse DNAs) PCR product from EC3/7C5 cells that is used in the method of the present invention.
XX	XX	Sequence 1434 BP; 336 A; 310 C; 333 G; 403 T; 52 other;
XX	XX	Query Match 35.58; Score 82; DB 25; Length 1434;
XX	XX	Best Local Similarity 71.96; Pred. No. 3.9e-10;
XX	XX	Matches 120; Conservative 0; Mismatches 40; Indels 7; Gaps
QY	59	GGCAGCTCTCATGTGCATGAAGGTTCACTGTCCCTAGTTCCCTTCCCCAGGAACCAACGACA

Db 904 GACATGTCATCTTTCAATGAAGGTTCAAGTGTCTAGTTCCTCCCTCCAGGCAAAANNACA 963  
QY 119 CGGGAGCTGCCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAGGGATGGTTTGTGT 171  
Db 964 CGGGAGCAGTCAAGGTTGCTCTGGGTAAAGCCTGTGAGCCTGGGAGCTAATCANNATAC 1023  
QY 172 AGGGCCCCATGTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218  
Db 1024 ATGGCTCTTTACCTACACTAGGGGATTTGACCTCTATCTCCACTC 1070

RESULT 11  
ID AAV04901 standard; DNA; 1400 BP.  
AC AAV04901;  
XX  
XX  
XX  
XX 30-APR-1998 (first entry)  
XX  
DE Nucleotide sequence of a portion of a neo-minichromosome.  
XX  
XX Neo-minichromosome; multicentric; dicentric; chromosome; euchromatin;  
KW heterochromatic DNA; minichromosome; artificial chromosome; organ;  
KW selectable marker; satellite artificial chromosome; SATAC; genomic locus;  
KW targeted integration; transgenic animal; therapeutic product;  
KW gene therapy; cloning vehicle; genomic DNA library; ss.  
XX  
OS Mus sp.  
XX  
XX WO9740183-A2.  
XX  
XX 30-OCT-1997.  
XX  
XX 10-APR-1997; 97WO-US05911.  
XX  
XX 07-AUG-1996; 96US-0695191.  
PR 10-APR-1996; 96US-0629822.  
PR 15-JUL-1996; 96US-0682080.  
XX  
XX (AMGE-) AMERICAN GENE THERAPY INC.  
PA (BIOL-) BIOLOGICAL RES CENT HUNGARIAN ACAD SCI.  
PA (UYLO-) UNIV LOMA LINDA.  
XX  
XX Hadlaczky G, Szalay AA;  
PI WPI; 1997-535860/49.  
XX  
XX Producing satellite artificial chromosomes or mini.chromosomes -  
PT useful for, e.g. cloning multiple proteins of a metabolic pathway or  
PT multivalent vaccines, etc.  
XX  
XX Claim 85; Page 204; 248pp; English.  
XX  
XX Nucleotide sequences AAV04900-902 are portions of a neo-minichromosome  
CC isolated from a mouse RNA gene. 3 products were obtained from a single  
CC amplification reaction, which suggests that the sequence of the DNA  
CC located between different sets of inverted repeats may differ.  
CC AAV04900-01 show high (96%) sequence homology to portions of DNA from  
CC intracisternal A-particles from mouse. The minichromosome is derived from  
CC multicentric, typically dicentric, chromosome that contains more  
CC euchromatin than heterochromatic DNA. The minichromosome is an example  
CC of an artificial chromosome. Artificial chromosomes can be produced by  
CC incorporating a DNA fragment comprising a selectable marker into a cells'  
CC genomic DNA, into or adjacent to an amplifiable region, and selecting a  
CC cell that comprises either a satellite artificial chromosome (SATAC)  
CC (see AAV04903-09), a de novo centromere, or a minichromosome (e.g.  
CC present sequence). Artificial chromosomes provide an extra genomic locus  
CC for targeted integration of megabase size DNA fragments that contain  
CC single or multiple genes. SATACs can be introduced into embryonic cells  
CC of non-human animals to produce transgenic animals that express a  
CC heterologous DNA encoding a therapeutic product, e.g. anti-HIV. The DNA  
CC may also encode antigens that induce immunoprotective response against

CC pathogen. These therapeutic products can be used in gene therapy. The  
CC artificial chromosomes are useful as cloning vehicles that accommodate  
CC entire genomes for the preparation of genomic DNA libraries, and also  
CC for the production of proteins which may be involved in a biochemical  
CC pathway or in multivalent vaccines.  
XX  
SQ Sequence 1400 BP; 341 A; 310 C; 345 G; 404 T; 0 other;  
Query Match 35.2%; Score 81.4; DB 18; Length 1400;  
Best Local Similarity 71.9%; Pred. No. 6.3e-18;  
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;  
QY 55 GGAAGGCACGTCGTGATTCGATGAAGTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAAAC 114  
Db 771 GGGAGACATGTCATCTTTCAAGAAGTTTGAGTGTCCCAAGTGTCTTCTCCAGGCAAAAC 830  
QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAGGGATGGTTT 167  
Db 831 GACACGGGAGCAGGTGCTCTGGGTAAAGCCCTGAGCCCTAAGAGCTAATCTCT 890  
QY 168 GTGTAGGCCCTTATGCTTTCACACTGGGGATCAGACCTCTACCTTCACCC 218  
Db 891 GTACATGGCTCTCTTTACCTACACACTGGGGATTTGACCTCTATCTCCACTC 941

RESULT 12  
ABX11083  
ID ABX11083 standard; DNA; 1400 BP.  
XX  
XX AC ABX11083;  
XX  
XX 24-APR-2003 (first entry)  
XX  
DE DNA sequence of neo-minichromosome PCR product #2 from EC3/7C5 cells.  
XX  
XX Nucleic acid amplification; artificial chromosome isolation; MAC;  
KW DNA delivery; mammalian artificial chromosome; gene therapy; organ;  
KW humanised genetically transformed animal; chromosomal element;  
KW gene product production system; transgenic; centromere function;  
KW information storage vehicle; artificial chromosome vector; human;  
KW species-specific artificial chromosome; mouse; bacteriophage lambda;  
KW bacteria; plasmid; neo-minichromosome; EC3/7C5 cell; mutant; ds.  
XX  
OS Homo sapiens.  
OS Mus sp.  
OS Bacteriophage lambda.  
OS Bacteriaceae.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FT misc\_feature 1..1400  
FT /tag= a  
FT /note= "All N nucleotides have been added by the  
FT index to match the number of bases this  
FT sequence is listed to contain in the  
FT sequence listing"

US2002160410-A1.  
31-OCT-2002.  
17-APR-2002; 2002US-0125767.  
28-NOV-2000; 2000US-0724693.  
PR 10-APR-1997; 97US-0835682.  
PR 10-APR-1996; 96US-0629822.  
PR 15-JUL-1996; 96US-0682080.  
PR 07-AUG-1996; 96US-0695191.  
XX  
XX (HADL/) HADLACZY G.  
PA (SZAL/) SZALAY A A.  
XX  
XX Hadlaczky G, Szalay AA;  
PI

```
XX DR WPI; 2003-265757/26.
XX PA Amplifying nucleic acid for constructing artificial chromosomes,
XX PT comprises introducing a DNA fragment into a cell to incorporate the DNA
XX PT into a chromosome, and identifying cells with chromosomes that has
XX PT undergone amplification
XX PS Example 2; Page 64; 107pp; English..
XX CC The present invention relates to methods for amplifying nucleic acids
XX CC in cells, for isolating artificial chromosomes, and preparing cell
XX CC lines that contain artificial chromosomes. The methods comprise
XX CC introducing a DNA fragment comprising a selectable marker into a cell,
XX CC growing the cell under selective conditions to produce cells that have
XX CC incorporated the DNA fragment or its portion into a chromosome, and
XX CC identifying from among the resulting cells those that include a
XX CC chromosome or its fragment with a portion that has undergone
XX CC amplification. The invention also discloses cells and cell lines that
XX CC contain the nucleic acids or artificial chromosomes. The methods of
XX CC the invention are useful for amplifying nucleic acids in cells, in
XX CC generating and isolating artificial chromosomes (e.g. mammalian
XX CC artificial chromosomes or MACs), and in delivering the chromosomes
XX CC to selected cells and tissues. The artificial chromosomes are useful
XX CC in gene therapy, gene product production systems, production
XX CC of humanised genetically transformed animal organs, production of
XX CC transgenic plants and animals that would employ chromosomal elements as
XX CC information storage vehicles, for analysis and study of centromere
XX CC function, for the production of artificial chromosome vectors, and for
XX CC the preparation of species-specific artificial chromosomes. The present
XX CC sequence represents a neo-minichromosome (composed of human,
XX CC bacteriophage lambda, plasmid, bacterial and mouse DNAs) PCR product
XX CC from EC37/7C5 cells that is used in the method of the present invention.
XX SS Sequence 1400 BP; 330 A; 299 C; 331 G; 388 T; 52 other;

Query Match 34.0%; Score 78.6; DB 25; Length 1400;
Best Local Similarity 69.6%; Pred. No. 5.9e-17;
Matches 119; Conservative 0; Mismatches 45; Indels 7; Gaps 1;

QY 55 GGAAGGACGCTGTCATTCATGAGTTCAAGTTCAGTGTCTTCCTCCCGAGGAAAAAC 114
DB 771 GGGAGACANNCTCATCTTTCAAGAGGTTGAGTGCCAAAGTGCTCTCTCCAGGCAAAAC 830

QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCCTAAGGATGGTTTT 167
DB 831 GACACGGGNNCAGGTCAGGTTGCTCTGGTAAAGCCTGTGAGCCTAAGAGCTAATCCT 890

QY 168 GTGTAGGGCCCTATGCTTGCACATCGGATCAGACCTCTACCTTACCC 218
DB 891 GTACATGNNCCTTTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 941

RESULT 13
AAQ85497
ID AAQ85497 standard; DNA; 50 BP.
XX AC AAQ85497;
XX DT 10-JAN-1996 (first entry)
XX DE DNA probe 3 detects DNA-protein complex in immortal cells.
XX KW DNA-protein complex; detection; proliferation; tumour formation;
XX KW diagnose; malignancy; biopsy; probe; ss.
XX OS Synthetic.
XX PS WO9502701-A1.
XX PD 26-JAN-1995.
XX PF 13-JUL-1994; 94WO-EP02307.
```

```
XX PR 15-JUL-1993; 93DE-4323727.
XX PA (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX PI Abken H, Albert W, Jungfer H, Abken HJ;
XX DR WPI; 1995-067344/09.
XX CC New DNA-protein complex characteristic of cells with unlimited
XX CC proliferation capacity - and its components and derived antibodies,
XX CC useful in diagnosing malignant tumours.
XX PS Claim 1; Page 51; 106pp; German.
XX CC The DNA shown is found in human or animal cells that have an unlimited
XX CC capacity for unlimited cell proliferation or tumour formation. They have
XX CC no ability to promote immortalisation of the cells, and are usually
XX CC found in a DNA-protein complex in the cell cytoplasm. The DNA is useful
XX CC for detection of these complexes, and diagnosis of malignant tumours.
XX CC Differentiation between cells with unlimited and only transient
XX CC proliferation is possible. (See also AAQ85493-54)
XX SS Sequence 50 BP; 13 A; 15 C; 9 G; 13 T; 0 other;

Query Match 15.2%; Score 35; DB 16; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
DB 1 GATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 35

RESULT 14
AAC89560/C
ID AAC89560 standard; DNA; 122186 BP.
XX AC AAC89560;
XX DT 08-MAR-2001 (first entry)
XX DE Human histone deacetylase HDAC-D coding sequence.
XX KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;
XX KW HDAC-D; cell cycle; tumourigenesis; cancer; inhibitor; antisense;
XX KW gene therapy; ds.
XX OS Homo sapiens.
XX PN WO200071703-A2.
XX PD 30-NOV-2000.
XX PF 03-MAY-2000; 2000WO-IB01252.
XX PR 03-MAY-1999; 99US-0132287.
XX PA (METH-) METHYLGENE INC.
XX PI Macleod AR, Li Z, Besterman JM;
XX DR WPI; 2001-016407/02.
XX CC Antisense oligonucleotide that inhibits expression of a histone
XX CC deacetylase, useful for treating and/or alleviating the symptoms of
XX CC neoplasia, or for inhibiting neoplastic cell growth in an animal -
XX PS Disclosure; Page 89-125; 125pp; English.
XX CC The present invention provides inhibitors of histone deacetylase enzymes
XX CC such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
XX CC inhibitors may be antisense strands or they may be compounds identified
```

CC by contacting the enzyme with the compound and measuring the resulting  
CC enzyme activity. These inhibitors are useful for treating cancers and for  
CC identifying which histone deacetylase is involved in a neoplasia.

[illegible]

**RESULT 15**

ABL81597/C

ID ABL81597 standard; cDNA; 475 bp.

XX  
AC

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:4575.

Human; ovarian cancer; ovarian tumour; cytostatic; qene; ss.

OS Homo sapiens.

PN WO200192581-A2.

AA  
PD  
06-DEC-2001.

PF 29-MAY-2001; 2001WO-US17756.

PR 26-MAY-2000; 2000US-207484P.

PA (CORI-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R;

DR WPI; 2002-122075/16.

Composition for therapy and diagnosis of ovarian cancer comprising  
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
 PT polypeptide, antibody specific to polypeptide or T cell expressing  
 PT polypeptide - polypeptide

PS Claim 1; SEQ ID 4575; 489pp; English.

The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (III) of an ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL7023 to ABL7934, (III) encoding (II) having a sequence (S2), a T cell population of (II), or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (IV) is detected preferably by polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising contacting T cells with (III) or (II). (III) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 21:30:48 ; Search time 1470 Seconds  
(without alignments)  
6428.656 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 gggggtcgagctaagca.....ttcaccatgaggtgtgtt 231

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.in.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.ov.\*

22: em.pat.\*

23: em.ph.\*

24: em.pl.\*

25: em.ro.\*

26: em.sts.\*

27: em.un.\*

28: em.vi.\*

29: em.htg\_hum.\*

30: em.htg\_inv.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_mam.\*

37: em.htg\_vrt.\*

38: em\_sy.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	231	100.0	1086	10	AF250998	AF250998 Mus muscu
2	231	100.0	1086	10	AF250999	AF250999 Mus muscu
3	231	100.0	2701	10	BC020078	BC020078 Mus muscu
4	231	100.0	42511	2	AC100548	AC100548 Mus muscu
5	231	100.0	64042	2	AC101394	AC101394 Mus muscu
6	231	100.0	66393	2	AC100424	AC100424 Mus muscu
7	231	100.0	66489	2	AC101278	AC101278 Mus muscu
8	231	100.0	66489	2	AC101278	AC101278 Mus muscu
9	231	100.0	70360	2	AC101122	AC101122 Mus muscu
10	231	100.0	71109	10	AL935271	AL935271 Mouse DNA
11	231	100.0	77127	10	BX294125	BX294125 Mouse DNA
12	231	100.0	84250	10	AF481949	AF481949 Mus muscu
13	231	100.0	98653	10	AL731664	AL731664 Mouse DNA
14	231	100.0	110000	2	AC132401	AC132401 Mus muscu
15	231	100.0	110000	2	AL929223_2	Continuation (3 of
16	231	100.0	132457	2	AC102180	AC102180 Mus muscu
17	231	100.0	134025	2	AC107697	AC107697 Mus muscu
18	231	100.0	134264	10	BX005191	BX005191 Mouse DNA
19	231	100.0	134717	10	AL626805	AL626805 Mouse DNA
20	231	100.0	141025	10	AL807250	AL807250 Mouse DNA
21	231	100.0	142902	2	AC119234	AC119234 Mus muscu
22	231	100.0	144387	10	AL844603	AL844603 Mouse DNA
23	231	100.0	149172	10	AL127278	AL127278 Mus muscu
24	231	100.0	150017	2	AL773509	AL773509 Mus muscu
25	231	100.0	155071	2	AC137713	AC137713 Mus muscu
26	231	100.0	156256	2	AC140843	AC140843 Mus muscu
27	231	100.0	161580	2	AC118267	AC118267 Mus muscu
28	231	100.0	162681	10	AC122059	AC122059 Mus muscu
29	231	100.0	164949	2	AC119839	AC119839 Mus muscu
30	231	100.0	165222	2	AC121307	AC121307 Mus muscu
31	231	100.0	165764	2	AC131767	AC131767 Mus muscu
32	231	100.0	166764	2	AC115875	AC115875 Mus muscu
33	231	100.0	166777	2	AC124106	AC124106 Mus muscu
34	231	100.0	167744	10	AC125537	AC125537 Mus muscu
35	231	100.0	171826	2	AC119810	AC119810 Mus muscu
36	231	100.0	172533	2	AC120551	AC120551 Mus muscu
37	231	100.0	173417	10	AL607143	AL607143 Mouse DNA
38	231	100.0	175217	10	AC079680	AC079680 Mus muscu
39	231	100.0	175448	2	AC138680	AC138680 Mus muscu
40	231	100.0	175692	2	AC139226	AC139226 Mus muscu
41	231	100.0	176184	2	AC119918	AC119918 Mus muscu
42	231	100.0	178821	10	AC122385	AC122385 Mus muscu
43	231	100.0	179068	10	AL627125	AL627125 Mouse DNA
44	231	100.0	179662	2	AC137557	AC137557 Mus muscu
45	231	100.0	179699	10	BX072541	BX072541 Mouse DNA

#### ALIGNMENTS

RESULT 1  
AF250998  
LOCUS AF250998 1086 bp DNA linear ROD 16-MAY-2001  
DEFINITION Mus musculus RTE-clone1 RNA transport element sequence.  
ACCESSION AF250998  
VERSION AF250998.1 GI:14090507  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 1086)  
AUTHORS Nappi, F., Schneider, R., Zolotukhin, A., Smulevitch, S.,  
Michalowski, D., Bear, J., Felber, B.K. and Pavlakis, G.N.  
TITLE Identification of a novel posttranscriptional regulatory element by

			/organism="Mus musculus"						
			/mol_type="genomic DNA"						
			/db_xref="taxon:10090"						
			/clone="RTE-clones"						
	misc_feature	381..627	/note=fragment M1; contains RNA transport element (RTE); posttranscriptional regulatory element"						
BASE COUNT	247 a	268 c	277 g	294 t					
ORIGIN									
	Query Match	100.0%; Score 231;	DB 10;	Length 1086;					
	Best local Similarity	100.0%;	Pred. No.	2e-65;					
Matches	231;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	GTGGGTCGCAGGCTAAGCACTGCACAGAGTAGTCTTGCGATCCTTTGGCATCCCTGTGAAGG	60						
Dd	393	GTGGGTCGCAGGCTAAGCACATGCACAGAGTAGTCTTGCGATCCTTTGGCATCCCTGTGAAGG	452						
QY	61	CAGCTCTGATTGCATGAAGGTTTCAGTGTCTTACTAGTTCCCTTCCCAGAAAAAACGACACG	120						
Dd	453	CAGCTCTGATTGCATGAAGGTTTCAGTGTCTTACTAGTTCCCTTCCCAGAAAAAACGACACG	512						
QY	121	GGAGCTGCCAGACACCCTCTCTGGGTGATGAGCCTTAAGGATGGTTTTGTGTAGGGCCCCCT	180						
Dd	513	GGAGCTGGCCAAGACCTCTCTGGGTGATGAGCCTTAAGGATGGTTTTGTGTAGGGCCCCCT	572						
QY	181	ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT	231						
Dd	573	ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT	623						
RESULT 3									
BC020078									
LOCUS	BC020078	2701 bp	mRNA	linear	ROD 20-SEP-2002				
DEFINITION	Mus musculus, similar to RIKEN CDNA I700066C05 gene, clone MGC:28125 IMAGE:3980327, mRNA, complete cds.								
ACCESSION	BC020078								
VERSION	BC020078.1	GI:	18043256						
KEYWORDS	MGC.								
SOURCE	Mus musculus	(house mouse)							
ORGANISM	Mus musculus								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
AUTHORS	1 (bases 1 to 2701) Straussberg,R. Direct Submission								
TITLE	Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA								
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov								
REMARK COMMENT	Contact: MGC help desk Email: cgabps@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cdDNA Library Preparation: Life Technologies, Inc. cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amgb@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.								
FEATURES SOURCE	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 36 Row: e Column: 4 This clone was selected for full length sequencing because it passed the following selection criteria: similarity but not identity to protein.								
FEATURES	Location/Qualifiers								
SOURCE	1.-2701								

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L15671  
Center clone name: 152\_Q\_17  
-----

\* NOTE: This record contains 54 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 699: contig of 699 bp in length  
700 799: gap of 100 bp  
800 1475: contig of 676 bp in length  
1476 1575: gap of 100 bp  
1576 2267: contig of 692 bp in length  
2268 2367: gap of 100 bp  
2368 3032: contig of 665 bp in length  
3033 3132: gap of 100 bp  
3133 3827: contig of 695 bp in length  
3828 3927: gap of 100 bp  
3928 4628: contig of 701 bp in length  
4629 4728: gap of 100 bp  
4729 5393: contig of 665 bp in length  
5394 5493: gap of 100 bp  
5494 6163: contig of 670 bp in length  
6164 6264: contig of 100 bp  
6264 6946: contig of 683 bp in length  
6947 7046: gap of 100 bp  
7047 7724: contig of 678 bp in length  
7725 7824: gap of 100 bp  
7825 8528: contig of 704 bp in length  
8529 8628: gap of 100 bp  
8629 9326: contig of 698 bp in length  
9327 9426: gap of 100 bp  
9427 10129: contig of 703 bp in length  
10130 10229: gap of 100 bp  
10230 10909: contig of 680 bp in length  
10910 11009: gap of 100 bp  
11010 11717: contig of 708 bp in length  
11718 11817: gap of 100 bp  
11818 12474: contig of 657 bp in length  
12475 12574: gap of 100 bp  
12575 13236: contig of 662 bp in length  
13237 13336: gap of 100 bp  
13337 14011: contig of 675 bp in length  
14012 14111: gap of 100 bp  
14112 14820: contig of 709 bp in length  
14821 14920: gap of 100 bp  
14921 15634: contig of 714 bp in length  
15635 15734: gap of 100 bp

```

15735 16412: contig of 678 bp in length
16413 16512: gap of 100 bp
16513 17218: contig of 706 bp in length
17219 18010: contig of 692 bp in length
18011 18110: gap of 100 bp
18111 18790: contig of 680 bp in length
18791 18890: gap of 100 bp
18891 19594: contig of 704 bp in length
19595 20382: contig of 688 bp in length
20383 20482: gap of 100 bp
20483 21185: contig of 683 bp in length
21186 21265: gap of 100 bp
21266 21972: contig of 707 bp in length
21973 22072: gap of 100 bp
22073 22753: contig of 681 bp in length
22754 22853: gap of 100 bp
22854 23527: contig of 674 bp in length
23528 23627: gap of 100 bp
23628 24334: contig of 707 bp in length
24335 25132: contig of 698 bp in length
25133 25232: gap of 100 bp
25233 25910: contig of 678 bp in length
25911 26010: gap of 100 bp
26011 26691: contig of 681 bp in length
26692 26791: gap of 100 bp
26792 27469: contig of 678 bp in length
27470 27569: gap of 100 bp
27570 28256: contig of 687 bp in length
28257 28356: gap of 100 bp
28357 29071: contig of 715 bp in length
29072 29171: gap of 100 bp
29172 29868: contig of 697 bp in length
29869 30719: contig of 751 bp in length
30720 30819: gap of 100 bp
30820 31527: contig of 708 bp in length
31528 32310: contig of 683 bp in length
32311 32410: gap of 100 bp
32411 33099: contig of 689 bp in length
33100 33199: gap of 100 bp
33200 33877: contig of 678 bp in length
33878 34666: contig of 689 bp in length
34667 35460: contig of 694 bp in length
35461 35560: gap of 100 bp
35561 36271: contig of 711 bp in length
36272 36371: gap of 100 bp
36372 37045: contig of 674 bp in length
37046 37145: gap of 100 bp
37146 37835: contig of 690 bp in length
37836 37935: gap of 100 bp
37936 38609: contig of 674 bp in length
38610 38709: gap of 100 bp
38710 39386: contig of 677 bp in length
39387 39486: gap of 100 bp
39487 40166: contig of 680 bp in length
40167 40266: gap of 100 bp
40267 40937: contig of 671 bp in length
40938 41037: gap of 100 bp
41039 41717: contig of 680 bp in length
41718 41817: gap of 100 bp
41818 42511: contig of 694 bp in length.

```

FEATURES

Location/Qualifiers  
1..42511  
/organism="Mus musculus"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-152017"

source

```

BASE COUNT 10831 a 7531 c 7714 g 10953 t 5482 others
ORIGIN
Query Match 100.0%; Score 231; DB 2; Length 42511;
Best Local Similarity 100.0%; Pred. NO. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATGCTGCTGTGGCATCTGTGGAAGG 60
|||||
Db 27042 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATGCTGCTGTGGCATCTGTGGAAGG 27101
|||||

QY 61 CAGCTCTGATTCATGATGAAGTTTCAGTGTCTAGTCTCCCTCCCGAGAGGAGACACG 120
|||||
Db 27102 CAGCTCTGATTCATGATGAAGTTTCAGTGTCTAGTCTCCCTCCCGAGAGGAGACACG 27161
|||||

QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGATGCTTTTGTAGGGCCCT 180
|||||
Db 27162 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGATGCTTTTGTAGGGCCCT 27221
|||||

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCATGAGGCTTGCCTT 231
|||||
Db 27222 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCATGAGGCTTGCCTT 27272

RESULT 5
AC101394 64042 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-119F14, LOW-PASS SEQUENCE SAMPLING.
AC101394
AC101394.1 GI:17060169
VERSION HTG; HTGS_PHASE0.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 64042)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE 2 (bases 1 to 64042)
Unpublished
AUTHORS Anderson,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faroo,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Liu,G.,
MacCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Matthews,C.,
McClean,C., Macdonald,P., Major,J., Marquis,N., Meltrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

```

Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L16611  
Center clone name: 119\_F\_14  
-----

\* NOTE: This record contains 79 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1  
\* 692: contig of 692 bp in length  
\* 792: gap of 100 bp  
\* 1511: contig of 719 bp in length  
\* 1611: gap of 100 bp  
\* 2313: contig of 702 bp in length  
\* 2413: gap of 100 bp  
\* 3108: contig of 695 bp in length  
\* 3208: gap of 100 bp  
\* 3209: contig of 721 bp in length  
\* 3930: gap of 100 bp  
\* 4030: gap of 100 bp  
\* 4745: contig of 716 bp in length  
\* 4845: gap of 100 bp  
\* 4846: gap of 100 bp  
\* 5559: contig of 714 bp in length  
\* 5560: gap of 100 bp  
\* 5660: gap of 100 bp  
\* 6404: contig of 745 bp in length  
\* 6405: gap of 100 bp  
\* 6505: gap of 100 bp  
\* 7235: contig of 731 bp in length  
\* 7335: gap of 100 bp  
\* 7336: gap of 100 bp  
\* 8068: contig of 733 bp in length  
\* 8168: gap of 100 bp  
\* 8875: contig of 707 bp in length  
\* 8876: gap of 100 bp  
\* 8976: gap of 100 bp  
\* 9697: contig of 722 bp in length  
\* 9797: gap of 100 bp  
\* 9798: gap of 100 bp  
\* 10508: contig of 711 bp in length  
\* 10608: gap of 100 bp  
\* 10609: gap of 100 bp  
\* 11313: contig of 705 bp in length  
\* 11314: gap of 100 bp  
\* 11413: gap of 100 bp  
\* 12111: contig of 698 bp in length  
\* 12211: gap of 100 bp  
\* 12212: gap of 100 bp  
\* 12926: contig of 715 bp in length  
\* 12927: gap of 100 bp  
\* 13027: gap of 100 bp  
\* 13733: contig of 707 bp in length  
\* 13734: gap of 100 bp  
\* 13834: gap of 100 bp  
\* 14544: contig of 711 bp in length  
\* 14644: gap of 100 bp  
\* 14645: gap of 100 bp  
\* 15345: contig of 701 bp in length  
\* 15346: gap of 100 bp  
\* 15445: gap of 100 bp  
\* 16168: contig of 723 bp in length  
\* 16268: gap of 100 bp  
\* 16269: gap of 100 bp  
\* 16970: contig of 702 bp in length  
\* 16971: gap of 100 bp  
\* 17070: gap of 100 bp  
\* 17764: contig of 694 bp in length  
\* 17765: gap of 100 bp  
\* 17865: gap of 100 bp  
\* 18570: contig of 706 bp in length  
\* 18571: gap of 100 bp  
\* 18671: gap of 100 bp  
\* 19389: contig of 719 bp in length  
\* 19489: gap of 100 bp  
\* 19490: gap of 100 bp  
\* 20197: contig of 708 bp in length  
\* 20297: gap of 100 bp  
\* 20298: gap of 100 bp  
\* 21030: contig of 733 bp in length  
\* 21031: gap of 100 bp  
\* 21130: gap of 100 bp  
\* 21859: contig of 729 bp in length  
\* 21959: gap of 100 bp  
\* 21960: gap of 100 bp  
\* 22675: contig of 716 bp in length  
\* 22775: gap of 100 bp  
\* 22776: gap of 100 bp  
\* 23496: contig of 721 bp in length

\* 23497: gap of 100 bp  
\* 23596: contig of 711 bp in length  
\* 24307: gap of 100 bp  
\* 24407: gap of 100 bp  
\* 25127: contig of 720 bp in length  
\* 25128: gap of 100 bp  
\* 25227: gap of 100 bp  
\* 25941: contig of 714 bp in length  
\* 26041: gap of 100 bp  
\* 26588: contig of 717 bp in length  
\* 26758: gap of 100 bp  
\* 26859: contig of 725 bp in length  
\* 27583: gap of 100 bp  
\* 27684: gap of 100 bp  
\* 28395: contig of 712 bp in length  
\* 28495: gap of 100 bp  
\* 29167: contig of 672 bp in length  
\* 29267: gap of 100 bp  
\* 29918: contig of 651 bp in length  
\* 30018: gap of 100 bp  
\* 30734: contig of 716 bp in length  
\* 30735: gap of 100 bp  
\* 30834: gap of 100 bp  
\* 31572: contig of 738 bp in length  
\* 31672: gap of 100 bp  
\* 31673: contig of 705 bp in length  
\* 32377: contig of 705 bp in length  
\* 32477: gap of 100 bp  
\* 33193: contig of 716 bp in length  
\* 33293: gap of 100 bp  
\* 33982: contig of 689 bp in length  
\* 34082: gap of 100 bp  
\* 34795: contig of 713 bp in length  
\* 34895: gap of 100 bp  
\* 34896: gap of 100 bp  
\* 35603: contig of 708 bp in length  
\* 35703: gap of 100 bp  
\* 36427: contig of 724 bp in length  
\* 36527: gap of 100 bp  
\* 37256: contig of 729 bp in length  
\* 37356: gap of 100 bp  
\* 38079: contig of 723 bp in length  
\* 38179: gap of 100 bp  
\* 38900: contig of 721 bp in length  
\* 39000: gap of 100 bp  
\* 39717: contig of 717 bp in length  
\* 39817: gap of 100 bp  
\* 40532: contig of 715 bp in length  
\* 40632: gap of 100 bp  
\* 41342: contig of 710 bp in length  
\* 41442: gap of 100 bp  
\* 42138: contig of 696 bp in length  
\* 42238: gap of 100 bp  
\* 42942: contig of 704 bp in length  
\* 43042: gap of 100 bp  
\* 43741: contig of 699 bp in length  
\* 43841: gap of 100 bp  
\* 44559: contig of 718 bp in length  
\* 44559: gap of 100 bp  
\* 44660: contig of 694 bp in length  
\* 45353: gap of 100 bp  
\* 45453: contig of 721 bp in length  
\* 46174: gap of 100 bp  
\* 46274: gap of 100 bp  
\* 46989: contig of 715 bp in length  
\* 47089: gap of 100 bp  
\* 47090: contig of 700 bp in length  
\* 47889: gap of 100 bp  
\* 47890: gap of 100 bp  
\* 48602: contig of 713 bp in length  
\* 48702: gap of 100 bp  
\* 49409: contig of 707 bp in length  
\* 49509: gap of 100 bp  
\* 50188: contig of 679 bp in length  
\* 50288: gap of 100 bp  
\* 51005: contig of 717 bp in length  
\* 51105: gap of 100 bp  
\* 51825: contig of 720 bp in length  
\* 51925: gap of 100 bp  
\* 52629: contig of 704 bp in length  
\* 52729: gap of 100 bp

```

* 52730 53448: contig of 719 bp in length
* 53449 53548: gap of 100 bp
* 54259 54259: contig of 711 bp in length
* 54260 54359: gap of 100 bp
* 54360 55074: contig of 715 bp in length
* 55075 55174: gap of 100 bp
* 55175 55883: contig of 709 bp in length
* 55884 55983: gap of 100 bp

Query Match      100.0%; Score 231; DB 2; Length 64042;
Best Local Similarity 100.0%; Pred. No. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGGAGCTAAGCACTGACAGAGATAGTCTGTGTCATCCCTGTGGAAGG 60
Db 32664 GTGGGTGGGAGCTAAGCACTGACAGAGATAGTCTGTGTCATCCCTGTGGAAGG 32723
QY 61 CAGCTCTGATTCATCAAGGTCAGTGCTCTAGTCCCTCCCGAGGAAACGACAG 120
Db 32724 CAGCTCTGATTCATCAAGGTCAGTGCTCTAGTCCCTCCCGAGGAAACGACAG 32783
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTAAGGATGTTTGTGTAGGGCCCT 180
Db 32784 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTAAGGATGTTTGTGTAGGGCCCT 32843
QY 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTACCCCATGAGGCTTGCTT 231
Db -32844 ATGCTTGACACATGGGGATCAGACCTCTACCTTACCCCATGAGGCTTGCTT 32894

- RESULT 6
AC100424/c
LOCUS      Mus musculus clone RP23-136F8, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC100424
VERSION    AC100424.1 GI:17047790
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 66393)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus, clone RP23-136F8
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 66393)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
            Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Kartas,A., Kells,C., LaRoque,K.,
            Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Triggilio,J., Vasiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15312
Center clone name: 136_F_8
-----

```

```

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

* 1 692: contig of 692 bp in length
* 693 792: gap of 100 bp
* 793 1509: contig of 717 bp in length
* 1510 1609: gap of 100 bp
* 1610 2315: contig of 706 bp in length
* 2316 2415: gap of 100 bp
* 2416 3132: contig of 717 bp in length
* 3133 3232: gap of 100 bp
* 3233 3312: contig of 680 bp in length
* 3313 4012: gap of 100 bp
* 4013 4682: contig of 670 bp in length
* 4683 5478: gap of 100 bp
* 5479 5577: contig of 695 bp in length
* 5578 6280: contig of 703 bp in length
* 6281 6380: gap of 100 bp
* 6381 7066: contig of 686 bp in length
* 7067 7166: gap of 100 bp
* 7167 7870: contig of 704 bp in length
* 7871 8655: contig of 695 bp in length
* 8656 8765: gap of 100 bp
* 8766 9472: contig of 707 bp in length
* 9473 9572: gap of 100 bp
* 9573 10230: contig of 658 bp in length
* 10231 10330: gap of 100 bp
* 10331 11024: contig of 694 bp in length
* 11025 11124: gap of 100 bp
* 11125 11809: contig of 685 bp in length
* 11810 11909: gap of 100 bp
* 11910 12589: contig of 680 bp in length
* 12590 12689: gap of 100 bp
* 12690 13374: contig of 685 bp in length
* 13375 13474: gap of 100 bp
* 13475 14170: contig of 696 bp in length
* 14171 14270: gap of 100 bp
* 14271 14967: contig of 697 bp in length
* 14968 15067: gap of 100 bp
* 15068 15770: contig of 703 bp in length
* 15771 15870: gap of 100 bp
* 15871 16588: contig of 718 bp in length
* 16589 17357: contig of 669 bp in length
* 17358 17457: gap of 100 bp
* 17458 18160: contig of 703 bp in length
* 18161 18260: gap of 100 bp
* 18261 18972: contig of 712 bp in length
* 18973 19072: gap of 100 bp
* 19073 19724: contig of 652 bp in length
* 19725 19824: gap of 100 bp
* 19825 20520: contig of 696 bp in length
* 20521 20620: gap of 100 bp

```

\* 20621 21326: contig of 706 bp in length  
\* 21327 21426: gap of 100 bp  
\* 21427 22117: contig of 691 bp in length  
\* 22118 22904: gap of 100 bp  
\* 22905 23004: contig of 687 bp in length  
\* 23005 23004: gap of 100 bp  
\* 23006 23805: contig of 701 bp in length  
\* 23806 24497: gap of 100 bp  
\* 24498 24597: contig of 692 bp in length  
\* 24598 25272: gap of 100 bp  
\* 25273 25372: contig of 675 bp in length  
\* 25373 26083: gap of 100 bp  
\* 26084 26183: contig of 711 bp in length  
\* 26184 26860: gap of 100 bp  
\* 26861 26960: contig of 677 bp in length  
\* 26961 27628: gap of 100 bp  
\* 27629 27728: contig of 668 bp in length  
\* 27729 28417: gap of 100 bp  
\* 28418 28517: contig of 689 bp in length  
\* 28518 29199: gap of 100 bp  
\* 29200 29299: contig of 682 bp in length  
\* 29300 29597: gap of 100 bp  
\* 29598 30057: contig of 658 bp in length  
\* 30058 30742: gap of 100 bp  
\* 30743 30842: contig of 685 bp in length  
\* 30843 31537: gap of 100 bp  
\* 31538 31637: contig of 695 bp in length  
\* 31638 32348: gap of 100 bp  
\* 32349 32448: contig of 711 bp in length  
\* 32449 33128: gap of 100 bp  
\* 33129 33228: contig of 680 bp in length  
\* 33229 33910: gap of 100 bp  
\* 33911 34010: contig of 682 bp in length  
\* 34011 34723: gap of 100 bp  
\* 34724 34823: contig of 713 bp in length  
\* 34824 35524: gap of 100 bp  
\* 35525 35624: contig of 701 bp in length  
\* 35625 36304: gap of 100 bp  
\* 36305 36404: contig of 680 bp in length  
\* 36405 37092: contig of 688 bp in length  
\* 37093 37192: gap of 100 bp  
\* 37193 37879: contig of 687 bp in length  
\* 37880 37979: gap of 100 bp  
\* 37980 38665: contig of 686 bp in length  
\* 38666 38765: gap of 100 bp  
\* 38766 39534: contig of 669 bp in length  
\* 39535 39634: gap of 100 bp  
\* 39635 40251: contig of 717 bp in length  
\* 40252 40351: gap of 100 bp  
\* 40352 41047: contig of 696 bp in length  
\* 41048 41147: gap of 100 bp  
\* 41148 41821: contig of 674 bp in length  
\* 41822 41921: gap of 100 bp  
\* 41922 42631: contig of 710 bp in length  
\* 42632 42731: gap of 100 bp  
\* 42732 43437: contig of 706 bp in length  
\* 43438 43537: gap of 100 bp  
\* 43538 44243: contig of 706 bp in length  
\* 44244 44343: gap of 100 bp  
\* 44344 45011: contig of 668 bp in length  
\* 45012 45111: gap of 100 bp  
\* 45112 45789: contig of 678 bp in length  
\* 45790 45889: gap of 100 bp  
\* 45890 46583: contig of 694 bp in length  
\* 46584 46683: gap of 100 bp  
\* 46684 47363: contig of 680 bp in length  
\* 47364 47463: gap of 100 bp  
\* 47464 48166: contig of 703 bp in length  
\* 48167 48266: gap of 100 bp  
\* 48267 48971: contig of 705 bp in length  
\* 48972 49071: gap of 100 bp  
\* 49072 49774: contig of 703 bp in length

\* 49775 49874: gap of 100 bp  
\* 49875 50562: contig of 688 bp in length  
\* 50563 50662: gap of 100 bp  
\* 50663 51336: contig of 674 bp in length  
\* 51337 51436: gap of 100 bp  
\* 51437 52136: contig of 700 bp in length  
\* 52137 52236: gap of 100 bp  
\* 52237 52967: contig of 731 bp in length  
\* 52968 53067: gap of 100 bp  
\* 53068 53748: contig of 681 bp in length  
\* 53749 53848: gap of 100 bp  
\* 53849 54539: contig of 691 bp in length  
\* 54540 54639: gap of 100 bp

Query Match 100.0%; Score 231; DB 2; Length 66393;  
Best Local Similarity 100.0%; Pred. No. 1.7e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGGTGGAGGCTAAGCACTGACAGAGGATAGCTTCTTCCCTCCCAAGGAAACACACACG 60  
DB 22021 GTGGGGTGGAGGCTAAGCACTGACAGAGGATAGCTTCTTCCCTCCCAAGGAAACACACACG 60  
QY 61 CACGCTGATTCATGAAGTTCAGTGTCTTCTTCCCTCCCAAGGAAACACACACG 120  
DB 21961 CACGCTGATTCATGAAGTTCAGTGTCTTCTTCCCTCCCAAGGAAACACACACG 120  
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTAAGGGATGCTTTGTGTAGGGCCCT 180  
DB 21901 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTAAGGGATGCTTTGTGTAGGGCCCT 180  
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231  
DB 21841 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 21791

RESULT 7  
AC101278 66489 bp DNA linear HTG 23-NOV-2001  
LOCUS Mus musculus clone RP23-101P16, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC101278  
ACCESSION AC101278  
VERSION AC101278.1 GI:17060053  
KEYWORDS HTG: HTGS-PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 66489)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE 1 (bases 1 to 66489)  
JOURNAL Mus musculus, clone RP23-101P16  
REFERENCE 2 (bases 1 to 66489)  
Unpublished  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L16354  
Center clone name: 10L\_P\_16  
-----

\* NOTE: This record contains 81 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
723: contig of 723 bp in length  
724: 823: gap of 100 bp  
824: 1542: contig of 719 bp in length  
1543: 1642: gap of 100 bp  
1643: 2341: contig of 699 bp in length  
2342: 2441: gap of 100 bp  
2442: 3174: contig of 733 bp in length  
3175: 3274: gap of 100 bp  
3275: 4014: contig of 740 bp in length  
4015: 4114: gap of 100 bp  
4115: 4846: contig of 732 bp in length  
4847: 4946: gap of 100 bp  
4947: 5656: contig of 710 bp in length  
5657: 5756: gap of 100 bp  
5757: 6473: contig of 717 bp in length  
6474: 6573: gap of 100 bp  
6574: 7317: contig of 744 bp in length  
7318: 7417: gap of 100 bp  
7418: 8129: contig of 712 bp in length  
8130: 8229: gap of 100 bp  
8230: 8953: contig of 724 bp in length  
8954: 9053: gap of 100 bp  
9054: 9838: contig of 785 bp in length  
9839: 9938: gap of 100 bp  
9939: 10667: contig of 729 bp in length  
10668: 10767: gap of 100 bp  
10768: 11503: contig of 736 bp in length  
11504: 11603: gap of 100 bp  
11604: 12346: contig of 743 bp in length  
12347: 12446: gap of 100 bp  
12447: 13168: contig of 722 bp in length  
13169: 13268: gap of 100 bp  
13269: 13995: contig of 727 bp in length  
13996: 14095: gap of 100 bp  
14096: 14810: contig of 715 bp in length  
14811: 14910: gap of 100 bp  
14911: 15587: contig of 677 bp in length  
15588: 16390: contig of 703 bp in length  
16391: 16490: gap of 100 bp  
16491: 17212: contig of 722 bp in length  
17213: 17312: gap of 100 bp  
17313: 18033: contig of 721 bp in length  
18034: 18133: gap of 100 bp  
18134: 18856: contig of 723 bp in length  
18857: 18956: gap of 100 bp  
18957: 19693: contig of 737 bp in length  
19694: 19794: gap of 100 bp  
19795: 20529: contig of 736 bp in length  
20530: 20629: gap of 100 bp  
20630: 21364: contig of 735 bp in length  
21365: 21464: gap of 100 bp  
21465: 22189: contig of 725 bp in length  
22190: 22289: gap of 100 bp  
22290: 22981: contig of 692 bp in length  
22982: 23081: gap of 100 bp  
23083: 23780: contig of 699 bp in length  
23781: 23880: gap of 100 bp  
23881: 24606: contig of 726 bp in length  
24607: 24706: gap of 100 bp  
24707: 25424: contig of 718 bp in length  
25425: 25524: gap of 100 bp  
25525: 26243: contig of 719 bp in length  
26244: 26343: gap of 100 bp  
26344: 27068: contig of 725 bp in length  
27069: 27168: gap of 100 bp  
27169: 27900: contig of 732 bp in length  
27901: 28000: gap of 100 bp  
28001: 28726: contig of 726 bp in length  
28727: 28826: gap of 100 bp  
28827: 29559: contig of 733 bp in length  
29560: 29659: gap of 100 bp  
29660: 30397: contig of 738 bp in length  
30398: 31212: contig of 715 bp in length  
31213: 31312: gap of 100 bp  
31313: 32025: contig of 713 bp in length  
32026: 32125: gap of 100 bp  
32126: 32855: contig of 730 bp in length  
32856: 32955: gap of 100 bp  
32956: 33674: contig of 719 bp in length  
33675: 33774: gap of 100 bp  
33775: 34480: contig of 706 bp in length  
34481: 34580: gap of 100 bp  
34581: 35294: contig of 714 bp in length  
35295: 35394: gap of 100 bp  
35395: 36121: contig of 727 bp in length  
36122: 36221: gap of 100 bp  
36222: 36952: contig of 731 bp in length  
36953: 37052: gap of 100 bp  
37053: 37771: contig of 719 bp in length  
37772: 37871: gap of 100 bp  
37872: 38601: contig of 730 bp in length  
38602: 38701: gap of 100 bp  
38702: 39431: contig of 730 bp in length  
39432: 39531: gap of 100 bp  
39532: 40256: contig of 725 bp in length  
40257: 40357: gap of 100 bp  
40357: 41067: contig of 711 bp in length  
41068: 41167: gap of 100 bp  
41168: 41897: contig of 730 bp in length  
41898: 41997: gap of 100 bp  
41998: 42712: contig of 715 bp in length  
42713: 42812: gap of 100 bp  
42813: 43533: contig of 721 bp in length  
43534: 43633: gap of 100 bp  
43634: 44354: contig of 721 bp in length  
44355: 44454: gap of 100 bp  
44455: 45187: contig of 733 bp in length  
45188: 46026: contig of 739 bp in length  
46027: 46126: gap of 100 bp  
46127: 46824: contig of 698 bp in length  
46825: 46924: gap of 100 bp  
46925: 47656: contig of 732 bp in length  
47657: 47756: gap of 100 bp  
47757: 48484: contig of 728 bp in length  
48485: 48584: gap of 100 bp  
48585: 49301: contig of 717 bp in length  
49302: 49401: gap of 100 bp



\* 49402 50119: contig of 718 bp in length  
\* 50120 50219: gap of 100 bp  
\* 50934 50934: contig of 715 bp in length  
\* 50935 51034: gap of 100 bp  
\* 51035 51766: contig of 732 bp in length  
\* 51767 51866: gap of 100 bp  
\* 51867 52575: contig of 709 bp in length  
\* 52576 52675: gap of 100 bp  
\* 52676 53400: contig of 725 bp in length  
\* 53401 53500: gap of 100 bp  
\* 53501 54217: contig of 717 bp in length  
\* 54218 54317: gap of 100 bp  
\* 54318 55003: contig of 686 bp in length  
\* 55004 55103: gap of 100 bp  
\* 55104 55811: contig of 708 bp in length  
\* 55812 55911: gap of 100 bp  
\* 55912 56634: contig of 723 bp in length  
\* 56635 56734: gap of 100 bp

Query Match 100.0%; Score 231; DB 2; Length 66489;  
Best Local Similarity 100.0%; Pred. No. 1.7e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGGCTGGAGCTAAGCACTGCACAGAGGATAGCTGTGTCATCCTGTGGAAGG 60  
Db 19918 GTGGGCTGGAGCTAAGCACTGCACAGAGGATAGCTGTGTCATCCTGTGGAAGG 19977  
QY 61 CAGCTGTGATGCATGAAGTTTCAGTGTCTTCCCTTCCCTCCAGGAAACACACACG 120  
Db 19978 CAGCTGTGATGCATGAAGTTTCAGTGTCTTCCCTTCCCTCCAGGAAACACACACG 20037  
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTAGGGCCCT 180  
Db 20038 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTAGGGCCCT 20097  
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCATGAGGCTGCTT 231  
Db 20098 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCATGAGGCTGCTT 20148

RESULT 8  
AC101278/c 66489 bp DNA linear HTG 23-NOV-2001  
LOCUS  
DEFINITION Mus musculus clone RP23-101P16, LOW-PASS SEQUENCE SAMPLING.  
AC101278  
VERSION AC101278.1 GI:17060053  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 66489)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
TITLE Mus musculus, clone RP23-101P16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 66489)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,  
Coote, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Glode, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,  
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McKean, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L16354  
Center clone name: 101\_P16  
-----  
\* NOTE: This record contains 81 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1  
\* 723: contig of 723 bp in length  
\* 724 823: gap of 100 bp  
\* 824 1542: contig of 719 bp in length  
\* 1543 1642: gap of 100 bp  
\* 1643 2341: contig of 699 bp in length  
\* 2342 2441: gap of 100 bp  
\* 2442 3174: contig of 733 bp in length  
\* 3175 3274: gap of 100 bp  
\* 3275 4014: contig of 740 bp in length  
\* 4015 4114: gap of 100 bp  
\* 4115 4846: contig of 732 bp in length  
\* 4847 4946: gap of 100 bp  
\* 4947 5656: contig of 710 bp in length  
\* 5657 5756: gap of 100 bp  
\* 5757 6473: contig of 717 bp in length  
\* 6474 6573: gap of 100 bp  
\* 6574 7317: contig of 744 bp in length  
\* 7318 7417: gap of 100 bp  
\* 7418 8129: contig of 712 bp in length  
\* 8130 8229: gap of 100 bp  
\* 8230 8953: contig of 724 bp in length  
\* 8954 9053: gap of 100 bp  
\* 9054 9838: contig of 785 bp in length  
\* 9839 10667: gap of 100 bp  
\* 10668 10767: contig of 729 bp in length  
\* 10768 11503: contig of 736 bp in length  
\* 11504 11603: gap of 100 bp  
\* 11604 12346: contig of 743 bp in length  
\* 12347 12446: gap of 100 bp  
\* 12447 13168: contig of 722 bp in length  
\* 13169 13268: gap of 100 bp  
\* 13269 13995: contig of 727 bp in length  
\* 13996 14095: gap of 100 bp  
\* 14096 14810: contig of 715 bp in length  
\* 14811 14910: gap of 100 bp  
\* 14911 15587: contig of 677 bp in length  
\* 15588 15687: gap of 100 bp  
\* 15688 16390: contig of 703 bp in length  
\* 16391 16490: gap of 100 bp  
\* 16491 17212: contig of 722 bp in length  
\* 17213 17312: gap of 100 bp

TITLE  
JOURNAL  
COMMENT

\* 17313 18033: contig of 721 bp in length  
\* 18034 18133: gap of 100 bp  
\* 18134 18856: contig of 723 bp in length  
\* 18857 18956: gap of 100 bp  
\* 18957 19693: contig of 737 bp in length  
\* 19694 19793: gap of 100 bp  
\* 19794 20529: contig of 736 bp in length  
\* 20530 20629: gap of 100 bp  
\* 20630 21364: contig of 735 bp in length  
\* 21365 21464: gap of 100 bp  
\* 21465 22189: contig of 725 bp in length  
\* 22190 22289: gap of 100 bp  
\* 22290 22981: contig of 692 bp in length  
\* 22982 23081: gap of 100 bp  
\* 23082 23780: contig of 699 bp in length  
\* 23781 23880: gap of 100 bp  
\* 23881 24606: contig of 726 bp in length  
\* 24607 24706: gap of 100 bp  
\* 24707 25424: contig of 718 bp in length  
\* 25425 25524: gap of 100 bp  
\* 25525 26243: contig of 719 bp in length  
\* 26244 26343: gap of 100 bp  
\* 26344 27068: contig of 725 bp in length  
\* 27069 27168: gap of 100 bp  
\* 27169 27900: contig of 732 bp in length  
\* 27901 28000: gap of 100 bp  
\* 28001 28726: contig of 726 bp in length  
\* 28727 28826: gap of 100 bp  
\* 28827 29559: contig of 733 bp in length  
\* 29560 29659: gap of 100 bp  
\* 29660 30397: contig of 738 bp in length  
\* 30398 30497: gap of 100 bp  
\* 30498 31212: contig of 715 bp in length  
\* 31213 31312: gap of 100 bp  
\* 31313 32025: contig of 713 bp in length  
\* 32026 32125: gap of 100 bp  
\* 32126 32855: contig of 730 bp in length  
\* 32856 32955: gap of 100 bp  
\* 32956 33674: contig of 719 bp in length  
\* 33675 33774: gap of 100 bp  
\* 33775 34480: contig of 706 bp in length  
\* 34481 34580: gap of 100 bp  
\* 34581 35294: contig of 714 bp in length  
\* 35295 35394: gap of 100 bp  
\* 35395 36121: contig of 727 bp in length  
\* 36122 36221: gap of 100 bp  
\* 36222 36952: contig of 731 bp in length  
\* 36953 37052: gap of 100 bp  
\* 37053 37771: contig of 719 bp in length  
\* 37772 37871: gap of 100 bp  
\* 37872 38601: contig of 730 bp in length  
\* 38602 38701: gap of 100 bp  
\* 38702 39431: contig of 730 bp in length  
\* 39432 39531: gap of 100 bp  
\* 39532 40257: contig of 725 bp in length  
\* 40257 40356: gap of 100 bp  
\* 40357 41067: contig of 711 bp in length  
\* 41068 41167: gap of 100 bp  
\* 41168 41897: contig of 730 bp in length  
\* 41898 41997: gap of 100 bp  
\* 41998 42712: contig of 715 bp in length  
\* 42713 42812: gap of 100 bp  
\* 42813 43533: contig of 721 bp in length  
\* 43534 43633: gap of 100 bp  
\* 43634 44354: contig of 721 bp in length  
\* 44355 44455: gap of 100 bp  
\* 44455 45187: contig of 733 bp in length  
\* 45188 45287: gap of 100 bp  
\* 45288 46026: contig of 739 bp in length  
\* 46027 46126: gap of 100 bp  
\* 46127 46824: contig of 698 bp in length  
\* 46825 46924: gap of 100 bp  
\* 46925 47656: contig of 732 bp in length

\* 47657 47756: gap of 100 bp  
\* 47757 48484: contig of 728 bp in length  
\* 48485 48584: gap of 100 bp  
\* 48585 49301: contig of 717 bp in length  
\* 49302 49401: gap of 100 bp  
\* 49402 50119: contig of 718 bp in length  
\* 50120 50219: gap of 100 bp  
\* 50220 50934: contig of 715 bp in length  
\* 50935 51034: gap of 100 bp  
\* 51035 51766: contig of 732 bp in length  
\* 51767 51866: gap of 100 bp  
\* 51867 52575: contig of 709 bp in length  
\* 52576 52675: gap of 100 bp  
\* 52676 53400: contig of 725 bp in length  
\* 53401 54217: gap of 100 bp  
\* 54218 54317: contig of 717 bp in length  
\* 54318 55003: contig of 686 bp in length  
\* 55004 55103: gap of 100 bp  
\* 55104 55811: contig of 708 bp in length  
\* 55812 55911: gap of 100 bp  
\* 55912 56634: contig of 723 bp in length  
\* 56635 56734: gap of 100 bp

Query Match 100.0%; Score 231; DB 2; Length 66489;  
Best Local Similarity 100.0%; Pred. No. 1.7e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTGCTGTTGGCATCTCTGGAAGG 60  
Db 44342 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTGCTGTTGGCATCTCTGGAAGG 44283  
Qy 61 CACGCTGATTCATGAAGTTCAGTCTCCTAGTTCCTCCCTCCAGAAAAAGACACG 120  
Db 44282 CACGCTGATTCATGAAGTTCAGTCTCCTAGTTCCTCCCTCCAGAAAAAGACACG 44223  
Qy 121 GGAGCTGGCAAGACCTCTCTGGGTGATGAGCCTAAGGATGTTGTGTAGGGCCCT 180  
Db 44222 GGAGCTGGCAAGACCTCTCTGGGTGATGAGCCTAAGGATGTTGTGTAGGGCCCT 44163  
Qy 181 ATGCTTCACACTGGGATCAGACCTCTACCTCACCATGAGGCTTGCCTT 231  
Db 44162 ATGCTTCACACTGGGATCAGACCTCTACCTCACCATGAGGCTTGCCTT 44112

RESULT 9  
AC101122/c 70360 bp DNA linear HTG 23-NOV-2001  
LOCUS Mus musculus clone RP23-93E18, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC101122  
ACCESSION AC101122  
VERSION AC101122.1 GI:17059896  
KEYWORDS HTG; HTGS, PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 70360)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
TITLE Mus musculus, clone RP23-93E18  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 70360)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choe, P., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farrow, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,  
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,  
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: LI4988

Center clone name: 93\_E\_18

-----

\* NOTE: This record contains 86 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 721: contig of 721 bp in length  
 722 821: gap of 100 bp  
 822 1528: contig of 707 bp in length  
 1529 1628: gap of 100 bp  
 1629 2332: contig of 704 bp in length  
 2333 2432: gap of 100 bp  
 2433 3144: contig of 712 bp in length  
 3145 3244: gap of 100 bp  
 3245 3967: contig of 723 bp in length  
 3968 4067: gap of 100 bp  
 4068 4791: contig of 724 bp in length  
 4792 4891: gap of 100 bp  
 4892 5599: contig of 708 bp in length  
 5600 5699: gap of 100 bp  
 5700 6425: contig of 726 bp in length  
 6426 6525: gap of 100 bp  
 6526 7219: contig of 694 bp in length  
 7220 7319: gap of 100 bp  
 7320 8037: contig of 718 bp in length  
 8038 8137: gap of 100 bp  
 8138 8846: contig of 709 bp in length  
 8847 8946: gap of 100 bp  
 8947 9664: contig of 718 bp in length  
 9665 9764: gap of 100 bp  
 9765 10491: contig of 727 bp in length  
 10492 10591: gap of 100 bp  
 10592 11316: contig of 725 bp in length  
 11317 11416: gap of 100 bp  
 11417 12126: contig of 710 bp in length  
 12127 12226: gap of 100 bp  
 12227 12943: contig of 717 bp in length  
 12944 13043: gap of 100 bp  
 13044 13755: contig of 712 bp in length  
 13756 13855: gap of 100 bp  
 13856 14559: contig of 704 bp in length  
 14560 14659: gap of 100 bp  
 14660 15404: contig of 745 bp in length

15405 15504: gap of 100 bp  
 15505 16228: contig of 724 bp in length  
 16229 16328: gap of 100 bp  
 16329 17033: contig of 705 bp in length  
 17034 17133: gap of 100 bp  
 17134 17938: contig of 705 bp in length  
 17939 17938: gap of 100 bp  
 18664 18664: contig of 726 bp in length  
 18665 18764: gap of 100 bp  
 18765 19475: contig of 711 bp in length  
 19476 19575: gap of 100 bp  
 19576 20305: contig of 730 bp in length  
 20306 20405: gap of 100 bp  
 20406 21141: contig of 736 bp in length  
 21142 21241: gap of 100 bp  
 21242 21965: contig of 724 bp in length  
 21966 22065: gap of 100 bp  
 22066 22804: contig of 739 bp in length  
 22805 22904: gap of 100 bp  
 22905 23637: contig of 733 bp in length  
 23638 23737: gap of 100 bp  
 23738 24472: contig of 735 bp in length  
 24473 24572: gap of 100 bp  
 24573 25300: contig of 728 bp in length  
 25301 25400: gap of 100 bp  
 25401 26122: contig of 722 bp in length  
 26123 26222: gap of 100 bp  
 26223 26935: contig of 713 bp in length  
 26936 27035: gap of 100 bp  
 27036 27770: contig of 735 bp in length  
 27771 27870: gap of 100 bp  
 27871 28550: contig of 680 bp in length  
 28551 28650: gap of 100 bp  
 28651 29358: contig of 708 bp in length  
 29359 29458: gap of 100 bp  
 29459 30183: contig of 725 bp in length  
 30184 30283: gap of 100 bp  
 30284 31012: contig of 729 bp in length  
 31013 31112: gap of 100 bp  
 31113 31829: contig of 717 bp in length  
 31830 31929: gap of 100 bp  
 31930 32642: contig of 713 bp in length  
 32643 32742: gap of 100 bp  
 32743 33463: contig of 721 bp in length  
 33464 33563: gap of 100 bp  
 33564 34280: contig of 717 bp in length  
 34281 34380: gap of 100 bp  
 34381 35089: contig of 709 bp in length  
 35090 35189: gap of 100 bp  
 35190 35922: contig of 733 bp in length  
 35923 36022: gap of 100 bp  
 36023 36739: contig of 717 bp in length  
 36740 36839: gap of 100 bp  
 36840 37552: contig of 713 bp in length  
 37553 37652: gap of 100 bp  
 37653 38375: contig of 723 bp in length  
 38376 38475: gap of 100 bp  
 38476 39195: contig of 720 bp in length  
 39196 39295: gap of 100 bp  
 39296 40019: contig of 724 bp in length  
 40020 40119: gap of 100 bp  
 40120 40844: contig of 725 bp in length  
 40845 40944: gap of 100 bp  
 40945 41660: contig of 716 bp in length  
 41661 41760: gap of 100 bp  
 41761 42445: contig of 685 bp in length  
 42446 42545: gap of 100 bp  
 42546 43241: contig of 696 bp in length  
 43242 43341: gap of 100 bp  
 43342 44062: contig of 721 bp in length  
 44063 44162: gap of 100 bp  
 44163 44878: contig of 716 bp in length  
 44879 44978: gap of 100 bp



1 (bases 1 to 77127)  
Wallis, J.  
Direct Submission  
Submitted (17-APR-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 18, 2003 this sequence version replaced gi:29500502.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information  
on the WormPEP database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP24-129F5 is  
from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

FEATURES  
source  
1. 77127  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP24-129F5"  
/clone\_lib="RPCI-24"  
BASE COUNT 21919 a 15230 c 15167 g 24811 t  
ORIGIN

Query Match 100.0%; Score 231; DB 10; Length 77127;  
Best Local Similarity 100.0%; Pred. No. 1.7e-65;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGCGCTAAGCACTGCACAGATAGCTGCTGGTGGCACTCTGTGGAGG 60  
|||||  
Db GTGGGTGGAGCGCTAAGCACTGCACAGATAGCTGCTGGTGGCACTCTGTGGAGG 33237  
|||||

QY 61 CAGCTGTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCCAGAAACGACACG 120  
|||||

Db 32326 CAGCTGTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCCAGAAACGACACG 32267  
|||||

QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGACGCTTAAGGGATGTTTGTGTAGGCCCT 180  
|||||

Db 32266 GGAGCTGGCCAGACCTCTCTGGGTGATGACGCTTAAGGGATGTTTGTGTAGGCCCT 32207  
|||||

QY 181 ATGCTTGACACTGGGATCAGACCTCTACCTTACCCATGAGCGTTCCT 231  
|||||

Db 32206 ATGCTTGACACTGGGATCAGACCTCTACCTTACCCATGAGCGTTCCT 32156  
|||||

RESULT 12  
AF481949/c 84250 bp DNA linear ROD 23-OCT-2002  
LOCUS Mus musculus DOM1 (Dom1) and DOM2 (Dom2) genes, exons 1 through 5  
DEFINITION

```

exon      65543..65571
/gene="Dom2"
/number=1
exon      71054..71685
/gene="Dom2"
/number=2
CDS       join(71058..71685,73128..73398,74282..74432,75229..75420)
/gene="Dom2"
/codon_start=1
/product="Dom2"
/protein_id="AAM47489.1"
/db_xref="GI:21322148"
/translation="MTPSISWGLLLAGLCLVPSELAEDVQETDTSQKQSPASHEI
ATNGLFALSIRELVHQNTNIFSPVSIATAFAMLSIGSKGDTHQILBGLQFNL
TGTSEADHISQHLQTLNRPDSEQLQSTGNGLFVNNDLKLVKFLPEAKNKAQEV
FVSNFAESEAKVINDFEKGTQKIVAEKVELQDQTFALANILFKGKKKPPDF
ENTFAEAFHVDKSTTVKVPMDLQSLGMLDVHHCSTLSSVLLMDYVGNASAVFLPEDG
KMOHLRQTSKELISKILLNRHRLVQIHPRLISGDYNLKLMSPLGITRIFNNGA
DLSGITEENAPLKLKSNVHKAVLTIDETGTEAANAIVFEAVPMSPPIIRFDHPFLFI
IFEHTQSPFVFGKVVDPDTHK"
/number=3
exon      74282..74432
/gene="Dom2"
/number=4
exon      75229..75506
/gene="Dom2"
/number=5
BASE COUNT 22271 a 19868 c 19190 g 22921 t
ORIGIN
Query Match      100.0%; Score 231; DB 10; Length 84250;
Best Local Similarity 100.0%; Pred. No. 1.6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGCTGGAGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 60
Db 8605 GTGGGCTGGAGCTAAGCACTGCACAGAGGATAGCTTGTGTCATCCTGTGGAAGG 8546
QY 61 CAGCTCTGATTGATGAAGCTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACACACAG 120
Db 8545 CAGCTCTGATTGATGAAGCTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACACACAG 8486
QY 121 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 180
Db 8485 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 8426
QY 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 8425 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 8375
RESULT 13
AL731664 98653 bp DNA linear ROD 24-JUN-2002
LOCUS Mouse DNA sequence from clone RP23-400M24 on chromosome 4, complete
DEFINITION sequence.
ACCESSION AL731664
VERSION AL731664
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 98653)
JOURNAL Lovell,J.
Direct Submission
COMMENT Submitted (22-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 26, 2002 this sequence version replaced gi:21540117.
----- Genome Center

```

```

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
FEATURES             Location/Qualifiers
     source            1..98653
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="4"
                     /clone="RP23-400M24"
                     /clone_lib="RPCI-23"
BASE COUNT 28064 a 22581 c 22820 g 25188 t
ORIGIN
Query Match      100.0%; Score 231; DB 10; Length 98653;
Best Local Similarity 100.0%; Pred. No. 1.6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGATAGTTCCTGTGTCATCCTGTGGAAGG 60
Db 43322 GTGGGTGGAGGCTAAGCACTGCACAGAGATAGTTCCTGTGTCATCCTGTGGAAGG 43381
QY 61 CAGCTCTGATTGATGAAGCTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACACACAG 120
Db 43382 CAGCTCTGATTGATGAAGCTTCAGTGTCTAGTTCCTTCCCTCCAGGAAACACACAG 43441
QY 121 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 180
Db 43442 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 43501
QY 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 43502 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 43552
RESULT 14
AC132401_0/c
WPCOMMENT
Sequence split into 5 fragments LOCUS AC132401 Accession AC132401
Fragment Name      Begin      End
AC132401_0         1        110000
AC132401_1        100001    210000
AC132401_2        200001    310000
AC132401_3        300001    410000
AC132401_4        400001    489146
LOCUS AC132401 489146 bp DNA linear HTG 03-SEP-2002
DEFINITION Mus musculus chromosome UNK clone RP23-40A8, WORKING DRAFT
SEQUENCE, 140 unordered pieces.
ACCESSION AC132401

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC132401.1 GI:22657894  
HTG: HTGS\_PHASE1: HTGS\_DRAFT.  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 489146)  
McPherson, J.D. and Waterston, R.H.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 489146)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Project Information -----  
Center project name: M\_BA0404A08  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-primer ET; 0% of reads  
Assembly: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 430636 bases at least Q40  
Consensus quality: 448443 bases at least Q30  
Consensus quality: 460211 bases at least Q20  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 140 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1212: contig of 1212 bp in length  
1213 1312: gap of unknown length  
1313 2777: contig of 1265 bp in length  
2578 2677: gap of unknown length  
2678 4509: contig of 1732 bp in length  
4510 4509: gap of unknown length  
4510 5592: contig of 1183 bp in length  
5593 5792: gap of unknown length  
5793 6999: contig of 1207 bp in length  
7000 7099: gap of unknown length  
7100 8301: contig of 1202 bp in length  
8302 8401: gap of unknown length  
8402 9700: contig of 1299 bp in length  
9701 9800: gap of unknown length  
9801 10916: contig of 1116 bp in length  
10917 11016: gap of unknown length  
11017 12583: contig of 1567 bp in length  
12584 12683: gap of unknown length  
12684 13755: contig of 1072 bp in length  
13756 13855: gap of unknown length  
13856 15575: contig of 1720 bp in length  
15576 15676: gap of unknown length  
15676 16894: contig of 1219 bp in length  
16895 16994: gap of unknown length  
16995 18323: contig of 1329 bp in length  
18324 18423: gap of unknown length  
18424 19659: contig of 1236 bp in length  
19660 19759: gap of unknown length  
19760 21346: contig of 1587 bp in length  
21347 21446: gap of unknown length  
21447 22722: contig of 1276 bp in length

22723 22822: gap of unknown length  
22823 23882: contig of 1060 bp in length  
23883 23982: gap of unknown length  
23983 25215: contig of 1233 bp in length  
25216 25316: gap of unknown length  
25316 26372: contig of 1057 bp in length  
26373 27911: gap of unknown length  
27911 28011: contig of 1439 bp in length  
28012 28938: contig of 1827 bp in length  
28939 29938: gap of unknown length  
29939 31555: contig of 1617 bp in length  
31556 33922: gap of unknown length  
33922 33992: contig of 1537 bp in length  
33993 34957: contig of 1565 bp in length  
34958 36321: gap of unknown length  
36321 38181: contig of 1364 bp in length  
38182 38282: contig of 1297 bp in length  
38283 39579: gap of unknown length  
39579 41309: contig of 1631 bp in length  
41310 42339: gap of unknown length  
42340 42639: contig of 1130 bp in length  
42640 44944: contig of 2305 bp in length  
44945 45044: gap of unknown length  
45045 46293: contig of 1249 bp in length  
46294 47882: gap of unknown length  
47883 49351: contig of 1489 bp in length  
49352 49451: gap of unknown length  
49452 50966: contig of 1515 bp in length  
50967 52365: gap of unknown length  
52366 52466: contig of 1299 bp in length  
52467 54068: gap of unknown length  
54069 55874: gap of unknown length  
55875 55974: contig of 1707 bp in length  
55975 57285: gap of unknown length  
57286 57385: contig of 1311 bp in length  
57386 59623: contig of 2238 bp in length  
59624 59723: gap of unknown length  
59724 61001: contig of 1278 bp in length  
61002 62734: gap of unknown length  
62735 62835: contig of 1633 bp in length  
62836 64266: gap of unknown length  
64267 64365: contig of 1431 bp in length  
64366 66070: gap of unknown length  
66071 66170: contig of 1705 bp in length  
66171 68449: gap of unknown length  
68450 68549: contig of 2279 bp in length  
68550 69975: gap of unknown length  
69976 70074: contig of 1425 bp in length  
70075 71869: gap of unknown length  
71870 71969: contig of 1794 bp in length  
71970 73384: gap of unknown length  
73385 73485: contig of 1416 bp in length  
73486 74799: gap of unknown length  
74800 74899: contig of 1315 bp in length  
74900 77399: gap of unknown length  
77400 77398: contig of 2499 bp in length  
77399 78516: gap of unknown length  
78517 78616: contig of 1018 bp in length  
78617 80110: gap of unknown length  
80111 80210: contig of 1494 bp in length  
80211 81382: gap of unknown length  
81383 81482: contig of 1172 bp in length  
81483 81482: gap of unknown length

